AIRS/AMSU/HSB Version 5 Level 2 Performance and Test Report

Edited by:
Thomas J. Hearty and Edward T. Olsen
Contributions by:
E. Fetzer, S-Y Lee, E. Fishbein, W.F. Irion, B. Kahn, E. Manning, S. Granger, B. Tian
Jet Propulsion Laboratory, California Institute of Technology

June, 2007
Version 3.1

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA
Submit Questions to:
http://airs.jpl.nasa.gov/AskAirs

Table of Contents

Table of Tables

Table of Figures

AIRS/AMSU/HSB Version 5 Level 2 Test Report

1. Executive Summary

This report describes tests that have been performed to ensure that the AIRS V5 Level 2 retrieval algorithms have been implemented correctly and that the data products have improved from V4. Most of the tests involve comparisons of V5 data products with ECMWF and/or V4 results and thus are not a validation. However, some analyses do include limited comparisons with satellite, radiosonde, laboratory, and in-situ data.

The V5 temperature retrieval algorithm produces an increased yield and improved error estimates from V4. The largest increase in yield is over the poles and the largest improvement in the quality of the retrievals is over land. Although the number of accepted retrievals over the poles increased from ~50% to ~90% the actual retrieved temperatures are similar in V4 and V5. Also, the number of temperature outliers is similar in both versions.

Although the V5 water vapor retrievals have a slight decrease in the yield from ~95% to ~85%, there are improvements in the quality of the accepted retrievals, improved error estimates, fewer outliers, and problems that were identified during the V4 validation were addressed. Specifically, there are no longer anomalously high water vapor retrievals over warm scenes and there is an improvement in the upper tropospheric dry bias and total water vapor wet bias.

The V5 Level 2 trace gas retrievals include Methane, a new unvalidated research product, and revised Carbon Monoxide and Ozone retrievals. The V5 Ozone retrieval uses more channels than V4 and an observationally based climatology rather than regression for a first guess. This results in V5 Ozone retrievals that are less biased in the mid to low troposphere. Also, the V5 Carbon Monoxide retrievals have improved sensitivity and accuracy over V4.0.9.

A comparison of the AIRS cloud top height with CloudSat shows good agreement with the AIRS cloud top height with slightly less agreement for multilayer clouds. Also, V5 may have a slight degradation in the sensitivity to detect low clouds over ocean.

In addition, the following parameters have been newly implemented or revised in V5: **Ptropopause**, **T_tropopause**, **GP_tropopause**, **olr**, **H2OMMRSat**. V5 also has improved emissivity retrievals.

The V5 algorithm introduced a linearly varying CO_2 concentration to address spurious temperature and water vapor trends identified in V4 that were thought to arise because of the incorrect CO_2 concentration used in the radiative transfer calculation, however, this implementation has had little effect to ameliorate these

issues. The spurious trends therefore probably have other contributing factors, possibly regression or calibration. As the baseline of AIRS observations continues to grow it will become more important to find the causes and implement solutions in V6 development.

2. Introduction

This AIRS V5 Level 2 test report describes tests that have been performed to assure that the AIRS mission achieves the primary objectives described in the AIRS Team Leader Proposal (ATLP): “to meet the climate research needs and the operational weather prediction requirements.” In addition, changes and improvements from V4 are described any liens in the standard retrieval products characterized. This document describes tests performed for the AIRS/AMSU and AIRS-Only processing streams and retrieval products. The AIRS/AMSU/HSB processing stream has not been tested for this report. Unless otherwise stated V4 refers to V4.0.9.0 and V5 refers to V5.0.14.0 (the official versions released to the Goddard Earth Sciences Data and Information Services Center [GES DISC]). For the cases where tests were performed with another version we have verified that there are no significant differences for that parameter in the version tested and the version released.

Since there is no previous release of an AIRS-Only retrieval algorithm, the tests have been performed only to assess the soundness of the current algorithm. The tests described here have two primary goals. First, they will determine if the retrieved products are meeting the performance specifications that have already been established in certain regions (e. g. over tropical oceans) for earlier versions using validation analyses. Second, these tests establish that retrieved quantities in unvalidated regions are ready for more refined validation analyses. These goals have been accomplished by checking for significant deviations from previous data releases, and by comparing with ECMWF, a few selected validation data sets, and other satellite measurements.

The measurement specifications for the AIRS Level 2 core products are described in Table 1. In addition to the core products listed in Table 1, the Level 2 standard product files include more than 100 related geophysical and quality assessment parameters. Additional testing has been performed on some parameters that are only available in the support and cloud cleared radiance files.

The specifications in Table 1 were defined prior to launch of the Aqua spacecraft. AIRS validation activities have been phased to ensure these specifications are met for increasingly difficult geophysical conditions. The AIRS Team Leader Proposal states that the V5 standard products will be validated for all non-polar conditions. That requirement partially motivates the testing planned here. However, scientific interest in polar regions with regard to the International Polar

Year (IPY) motivates additional tests for very cold conditions. Section 2 of this report broadly defines the geophysical conditions for these tests.

In order for the AIRS data products to be useful for long-term climate trending, they must be stable and have minimal calibration, validation, or processing anomalies. Therefore the core products have been examined for "Focus Days" (FD) from 2002-2005 to verify the stability of the data set.

Section 3 describes the tests that have been performed for all relevant parameters in the AIRS standard, support, and cloud cleared radiance products. Section 4 describes non-physical "climatologies" that can sometimes be produced by the AIRS retrieval and elaborates on the non-physical retrievals found in each version (starting with V4). Section 5 lists the major changes that have gone into the development of the V5 PGE.

Table 1 Uncertainty and Resolution of AIRS/AMSU/HSB Retrieved Core Products

Core Product	RMS Uncertainty	Vertical Resolution	Horizontal Resolution
Cloud-Clear IR Radiance	1 K	N/A	45 x 45 km
Sea Surface Temperature	0.5 K	N/A	45 x 45 km
Land Surface Temperature	1 K	N/A	45 x 45 km
Temperature Profile	1 K	1 km below 700 mb 2 km 30-700 mb	45 x 45 km
Humidity Profile	15%	2 km in troposphere	45 x 45 km
Total Precipitable Water	5%	N/A	45 x 45 km
Fractional Cloud Cover	5%	N/A	45 x 45 km
Cloud Top Height	0.5 km	N/A	45 x 45 km
Cloud Top Temperature	1 K	N/A	45 x 45 km

3. Geophysical Conditions of the Tests

The AIRS products are being validated for increasingly complex geophysical conditions. The testing described in this plan will help establish that the performance of earlier data releases under less complex conditions is met for the V5, and, that the next round of validation analyses can be performed for more difficult conditions. For the purposes of this and later tests we define four broad climate states: non-frozen ocean, warm-season land, cold-season land, and polar winter. The broad conditions comprising these states and their current validation status are described here:

- **Non-frozen Ocean** -- oceanic regions without ice cover. The performance for AIRS is well established for these conditions for retrieved products from both v3 (Fetzer et al. 2003; Gettelman et al. 2004; Fetzer et al. 2004) and V4 (McMillan et al. 2005; Fetzer et al. 2005; see Fetzer 2006 for an overview and references; Kahn et al. 2006, 2006a).
- **Warm-Season Land** – ice-free summertime land at all latitudes, and all tropical and subtropical land areas. Fetzer et al. (2005), Divakarla et al. (2006), Tobin et al. (2006) and Susskind et al. (2006) describe results for warm-season land. Note that the AIRS retrievals for these conditions typically do not meet the specification stated in Table 1 in the bottom 1-2 km. Note also that Section 4 of this plan singles out deserts as particularly challenging to the AIRS retrieval system.
- **Cold-season land** – wintertime land between ±60 degrees with mean temperatures below freezing, ice-covered ocean equatorward of ±60 degrees latitude, and summertime over permanent ice caps. The only validation studies to date for these Antarctic conditions are Gettelman et al. (2006) and Ye et al. (2007).
- **Polar ice** – cold-season land and ice-covered ocean poleward of ±60 degrees latitude. Informal studies have shown these conditions to be challenging for the AIRS retrieval algorithms; see Section 4 of this report.

The four climate states are illustrated in Figure 1 for January 22, 2004. Seawater is considered frozen if the surface temperature in ECMWF is below 271.3 K for water and 273.0 K for land.

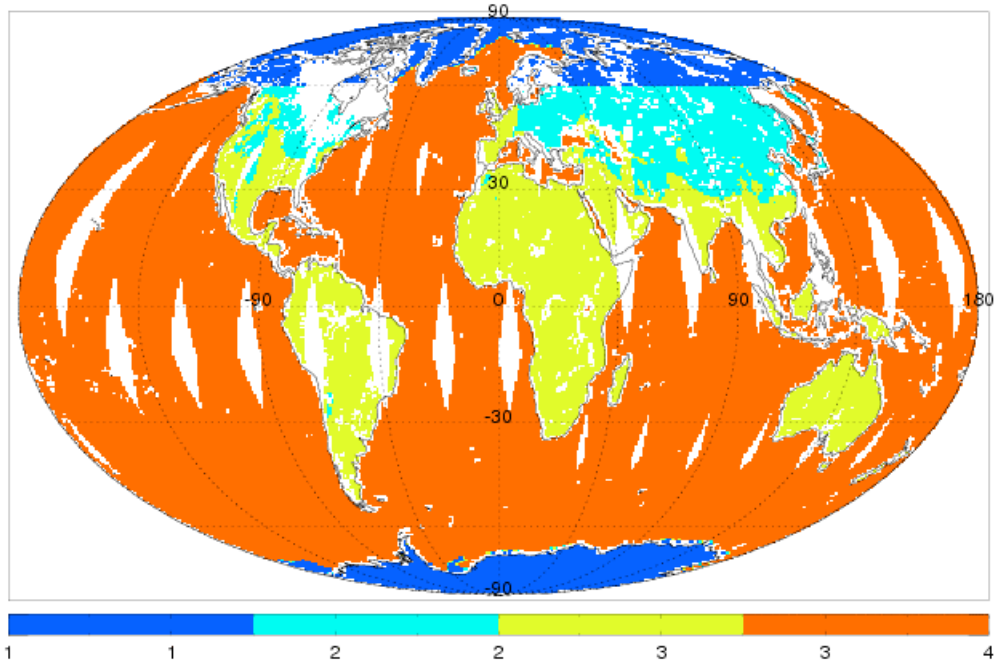

Figure 1. The geographic region definitions are displayed for January 22, 2004. Orange is "non-frozen ocean," Yellow is "warm-season land," light blue is "cold-season land," and dark blue is "polar winter."

4. Test Results

4.1. Limit Tests

The fundamental requirement for all of the parameters is that they must fall within the expected ranges. Most of these limit tests have been performed as part of the AIRS System Integration and Test Report (SITR).

4.2. Atmospheric Temperature

Parameter List

- **TAirSup**
- **TAirStd**
- **TAirStdErr**
- **Qual_Temp_Profile_Top**
- **Qual_Temp_Profile_Mid**
- **Qual_Temp_Profile_Bot**
- **TSurfAir**
- **TAirMWOnlyStd**
- **Qual_MW_Only_Temp_Strat**
- **Qual_MW_Only_Temp_Tropo**

The AIRS temperature retrieval quality control parameterization partitions the atmosphere into 3 slabs. The boundaries of the temperature quality slabs are given by **Press_mid_top_bndry** and **Press_bot_mid_bndry**, where **nStd_mid_top_bndry** and **nStd_bot_mid_bndry** are the nearest standard levels (counting from 1 at the surface) to these boundaries. The parameters **nSup_mid_top_bndry** and **nSup_bot_mid_bndry** are the nearest support level boundaries (counting from 1 at the top of the atmosphere). **Press_valid_bottom** (the bottom pressure at which temperature, water vapor, and ozone profiles are valid) from V4 has been replaced by **PBest** and **PGood** indicating the lowest pressure level where the data quality flags are 0 (**PBest**) or 0 or 1 (**PGood**). Quality flags set to 2 (do not use) are excluded from the analysis in this report. Since the broad temperature layer definitions have no data to which they can be compared we simply insured that the reported values are consistent with each other, within physically reasonable values for atmospheric pressures, and that the yields increase toward higher altitudes. These requirements are consistent with the yields displayed in Figure 2. Refer to the Documents:

V5_L2_Standard_Product_QuickStart.pdf

V5_L2_Quality_Control_and_Error_Estimation.pdf

The 100 level atmospheric temperature parameter, **TAirSup**, provided in the Level 2 support files has been converted to ~1-2 km thick layers and compared to similarly computed layer quantities derived from ECMWF forecast data. Figure 2 shows that the V5 PGE is meeting the RMS requirements from the ATLP in the free troposphere for the **Qual_Temp_Profile_*** = 0 cases but the RMS values are slightly larger in the boundary layer and above the tropopause. Also, the bias and RMS in V5 has less of an oscillatory structure than in V4. The yield in V5 has increased from V4 in all atmospheric layers except from ~100-200 mb. All three versions show a slight increasing trend in the bias at altitudes below ~800 mb. This trend was first pointed out by Divakarla et al. (2006) and will be discussed further in Section 4.14.

Globally V5 has a lower bias and RMS with respect to ECMWF in most atmospheric layers. Figure 3 divides the retrieval bias and RMS for Qual_* = 0 and 1 cases into the 4 broad climatological regimes described in Section 0. The largest gains in yield in V5 are over frozen surfaces. The largest improvement in the bias with respect to ECMWF is over warm season land.

Figure 4 shows global maps of the temperature bias at 850 for 2 Focus Days processed with V4 and V5. V5 has increased yield over land, poles, and sunglint regions over ocean. However, some regions with increased yield also have a larger bias. This should be investigated further in the V5 validation campaign.

The microwave only temperature retrievals (**TAirMWOnlyStd**) have not changed significantly since V4 (see, Figure 5).

Error Assessment Testing

Figure 6 displays the reduced χ^2 for the atmospheric temperature layers where,

$$N \times \chi^2 = \sum_{i=1}^{N} \left(\frac{T_{AIRS} - T_{ECMWF}}{T_{err}} \right)^2 . \qquad (1)$$

The reduced χ^2s in V5 are much smaller than those in V4 at altitudes below ~100 mb. This indicates either (1) better agreement with ECMWF or (2) improved error estimates, or both. Since the bias over ocean is similar in V4 and V5 and the V5 χ^2s are smaller, the smaller reduced χ^2s are due to the improved error estimate.

Outlier Testing

We define outliers as in the NIST/SEMATECH e-Handbook of Statistical Methods (http://www-09.nist.gov/div898/handbook/prc/section1/prc16.htm) using the InterQuartile Difference (IQD; the difference between the lower and the upper quartiles). Mild outliers are those that are 1.5 × IQD below the lower quartile or above the upper quartiles. Extreme outliers are those 3 × IQD below the lower quartiles or above the upper quartile. Figure 7 shows that the number of mild and extreme outliers is similar in V4 and V5 processed with AMSU. The AIRS-Only processing has significantly more outliers in the lower troposphere than the AIRS/AMSU retrievals.

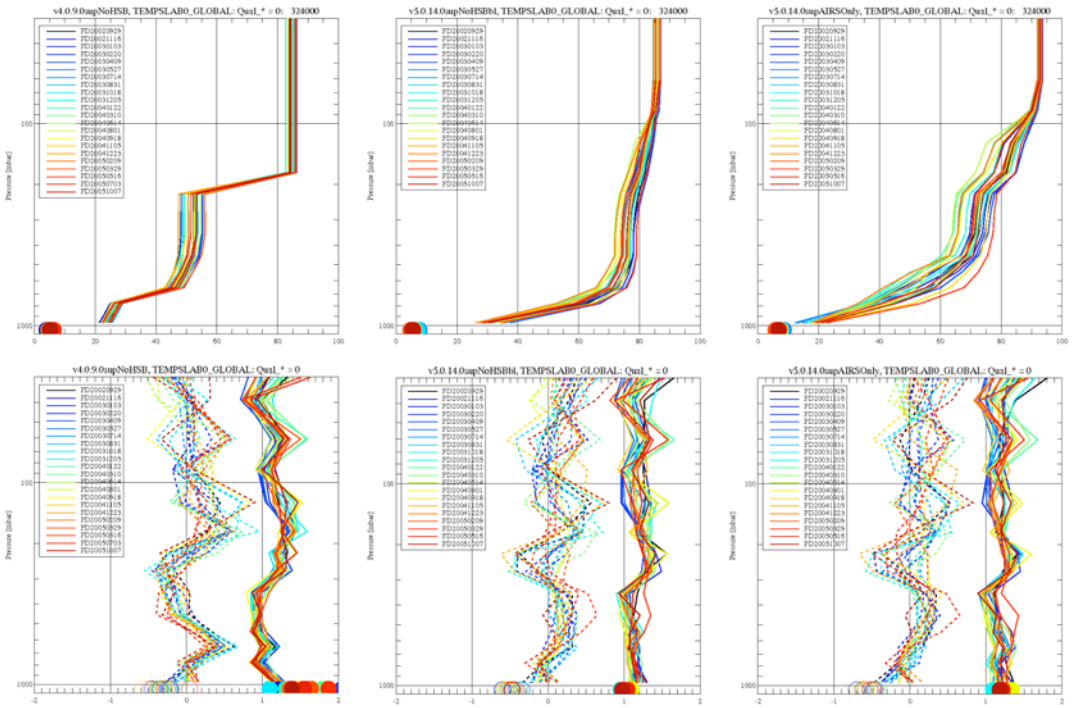

Figure 2 The top panels show the yield for atmospheric temperature retrievals with Qual_* = 0 for Focus days processed with V4.0.9.0, V5.0.11.0 AIRS/AMSU and V5.0.7.0 AIRS-Only (from left to right). The filled circles at the bottom of the plots show the surface temperature yield for Qual_Surf = 0. The bottom panels show the bias and RMS difference with respect to ECMWF for the same focus days. The open and filled circles at the bottom of the figures show the bias and RMS of the surface temperature retrievals.

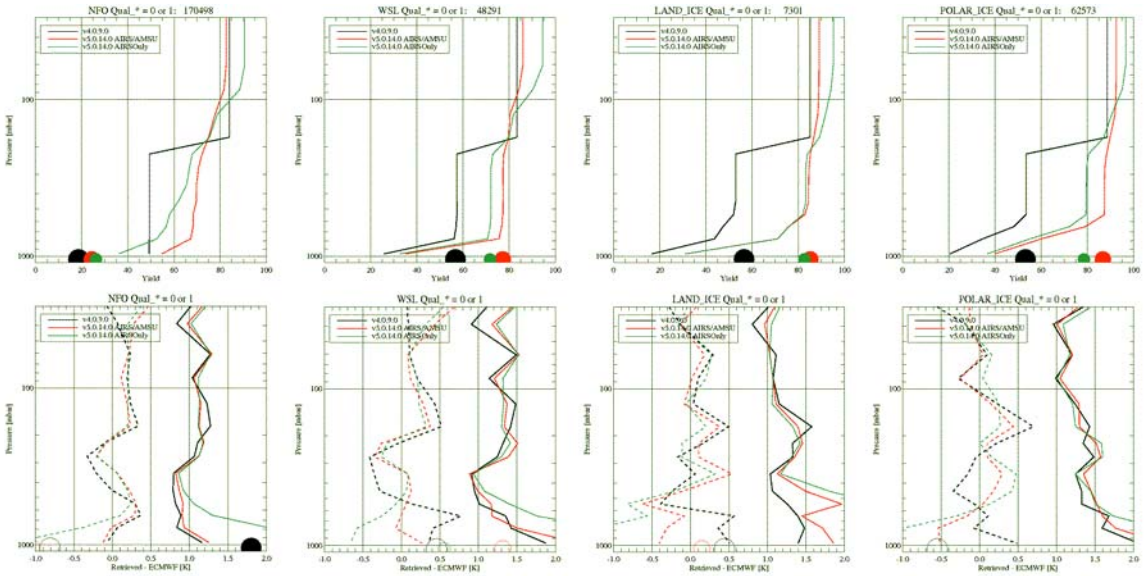

Figure 3 The top panels show the yield averaged over the Focus Days shown in Figure 2 for Qual_* = 0 and 1 cases in the four different climatological regimes described in Section 0 (NFO, WSL, Land Ice, Polar Ice from left to right). The bottom panels show the average bias and RMS for the same climatological regimes.

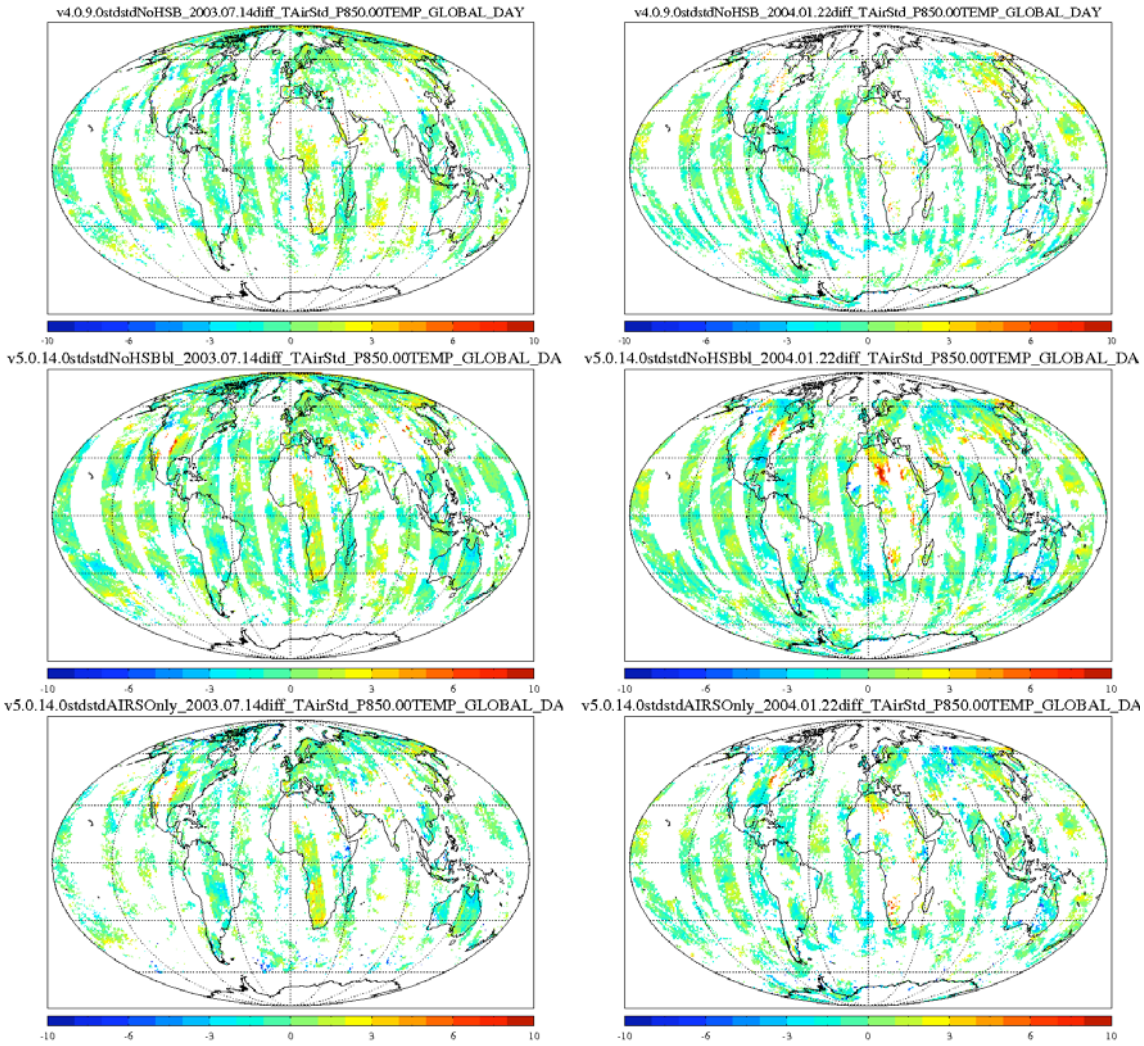

Figure 4 The difference between the AIRS retrieved temperature and ECMWF at 850 mb for 2 days using V4.0.9.0 (top), V5.0.14.0 AIRS/AMSU (middle), and V5.0.14.0 AIRS-Only (bottom).

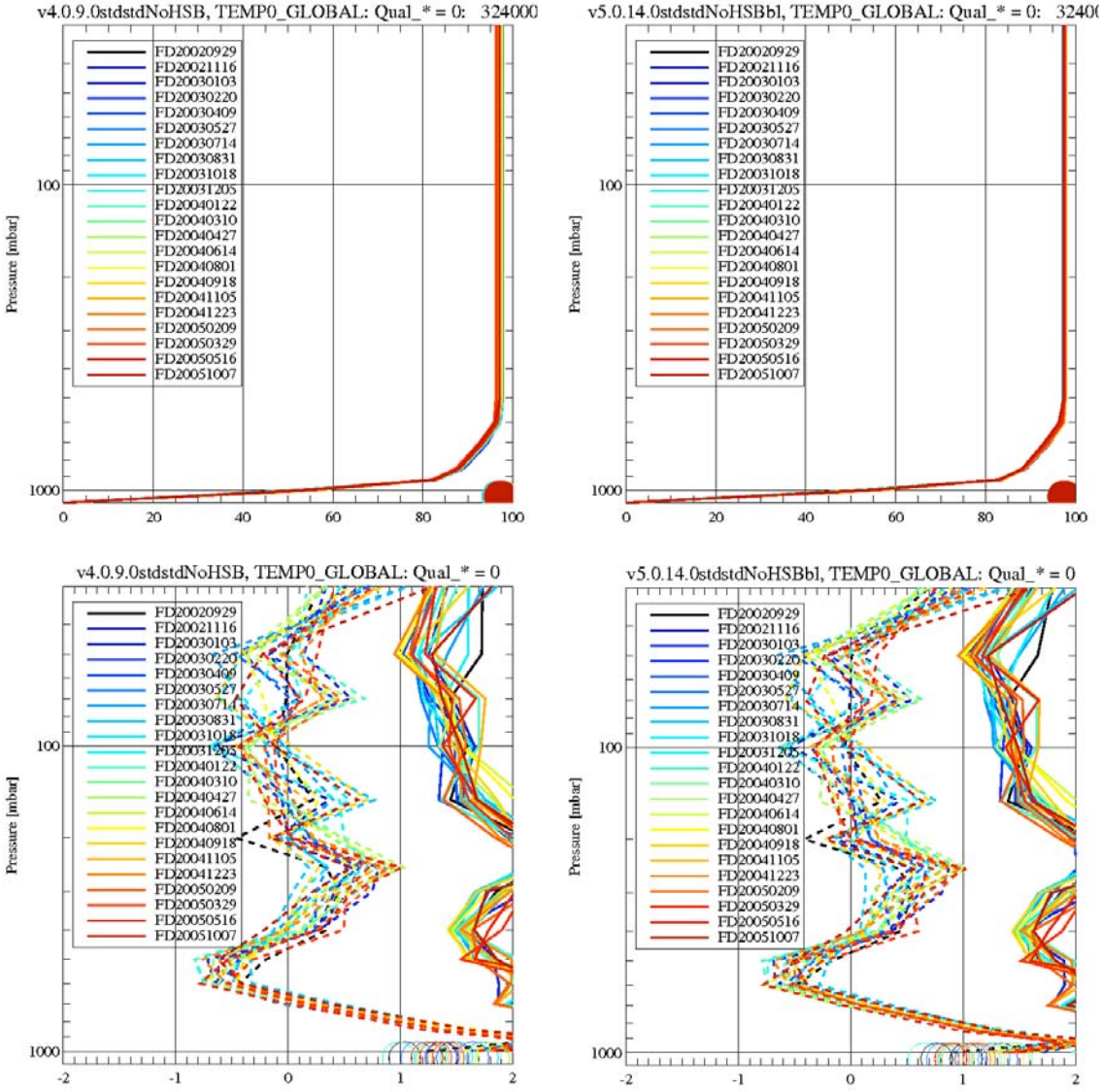

Figure 5 The top panels show the yield for Qual_MW_Only_Temp_Tropo and Qual_MW_Only_Temp_Strat is nearly 100% globally. The bottom panels show the bias (dashed lines) and RMS (sold lines) with respect to ECMWF. The bias and RMS of TAirMWOnlyStd have not changed significantly from V4.

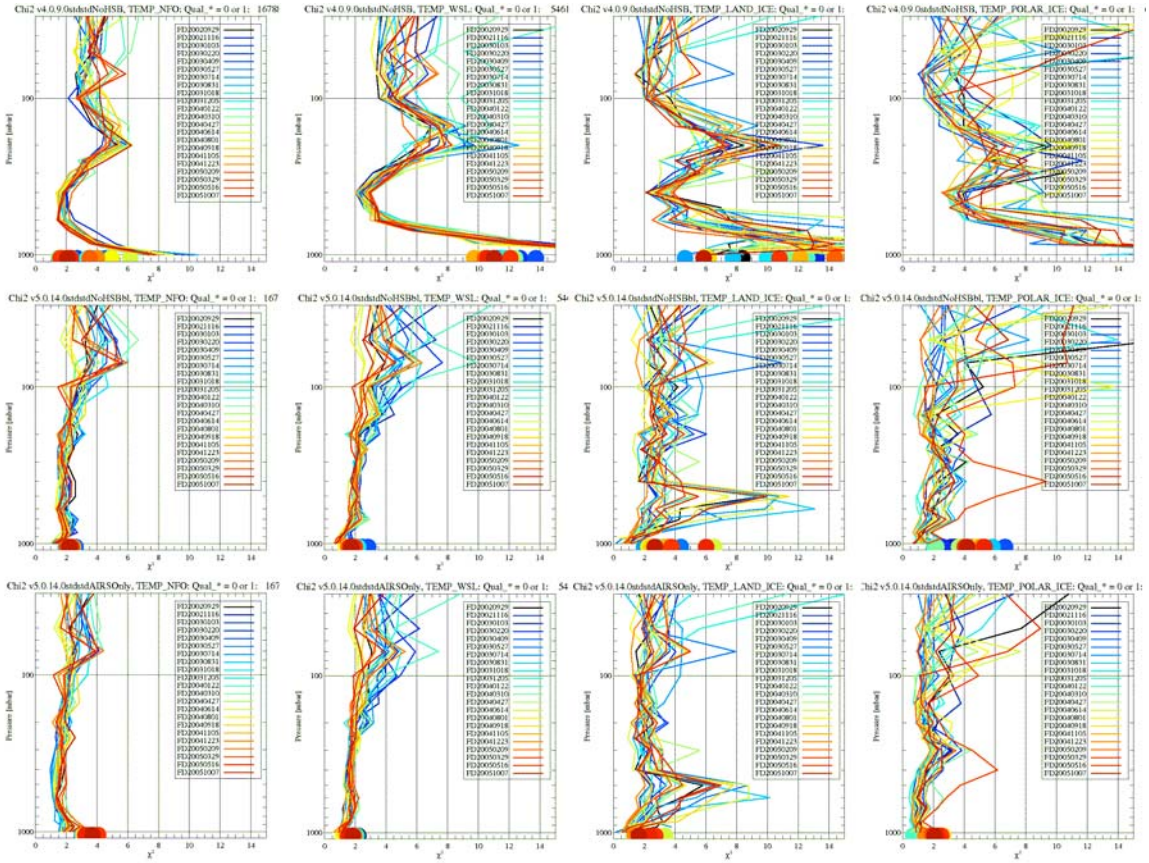

Figure 6 The reduced χ^2 of the atmospheric and surface temperature parameters are displayed for the 4 different geophysical regimes (from left to right they are NFO, WSL, Frozen land, Polar Ice) examined in this report. From top to bottom the figures are for V4, V5 AIRS/AMSU, V5 AIRS-Only.

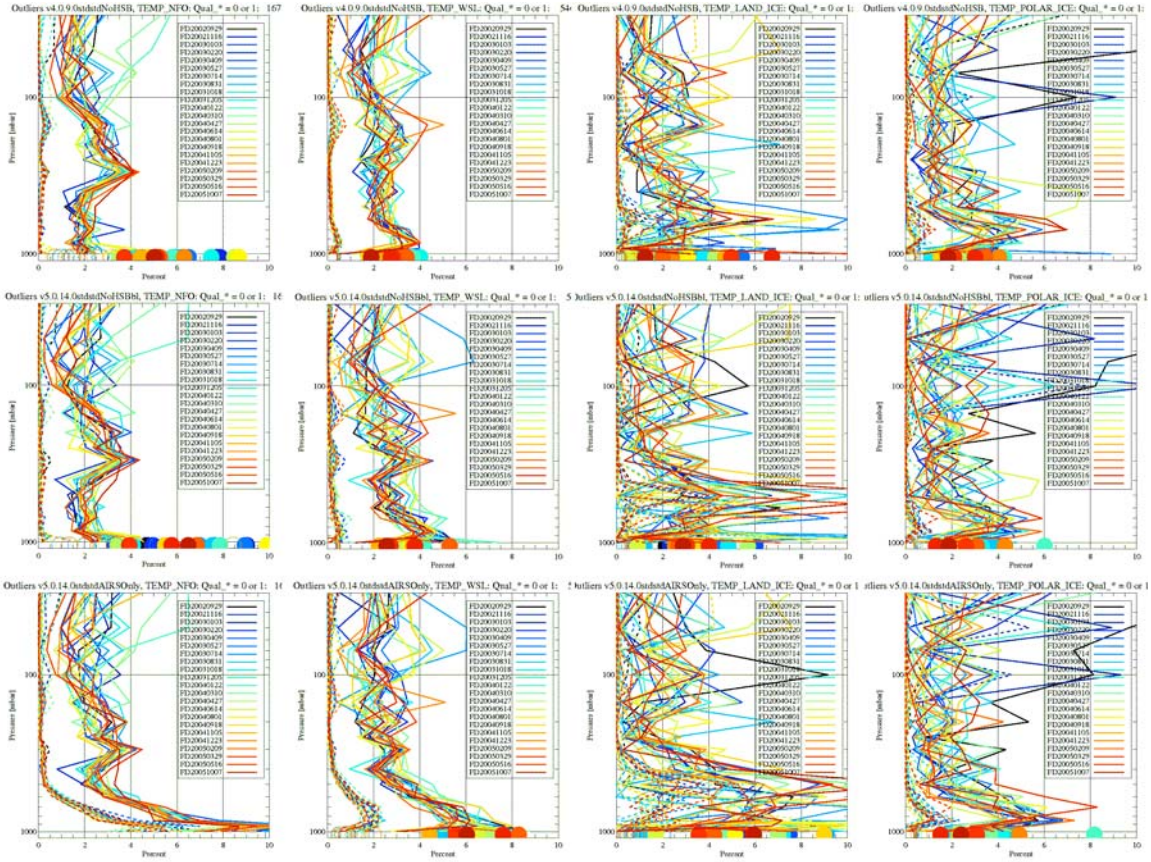

Figure 7 The percent of "mild outliers" and "extreme outliers" is displayed for the 4 different geophysical regimes (from left to right they are NFO, WSL, Frozen land, Polar Ice) examined in this report. The figure shows that over non-frozen ocean V5 has slightly fewer outliers than V4 but over land and ice covered surfaces there is a slight increase in outliers near the surface. The solid lines indicate inner-fence outliers and the dashed lines indicate outerfence outliers. The circles at the bottom indicate the number of surface temperature outliers.

4.3. Surface Temperature

Parameter List

- **TSurfStd**
- **TSurfStdErr**
- **Qual_Surf**
- **TSurfAirStd**
- **TSurfAirStdErr**

The AIRS Team Leader Proposal requires that the surface temperature have an RMS uncertainty of ~0.5 K over ocean and ~1 K over land. Since ECMWF is not sufficiently accurate to test this requirement over land or ice, these tests have been performed by comparing them to in-situ surface air temperature measurements obtained at Dome C in Antarctica. However, since the measurements at Dome C only sample a small subset of the AIRS data, comparisons have also been performed with ECMWF over land and ice.

Comparison with ECMWF

The yield for Qual_Surf = 0 is similar in V4 and V5 (Figure 2). Although the yield for Qual_Surf = 0 cases is very small < 10%, the RMS difference in V5 is much better than that of V4 and nearly meeting the requirements in the ATLP. Figure 8 shows that the bias in temperature over Non-Frozen Ocean is different between day and night. This is due to ECMWF measuring the “bucket temperature” while AIRS measures the skin temperature that has a larger diurnal variation.

Additional observations

- The yield for Qual_Surf = 0 and 1 cases increased in V5 for all surface types (Figure 3), however, the bias and RMS in V5 are larger than they were in V4 (Figure 9).
- Figure 9 shows that for Warm Season Land and Polar Ice the RMS differences with respect to ECMWF are much larger than over Non-Frozen Ocean.
- Figure 7shows that V5 has a similar percentage of outliers as V4 for the SST over all surface types.
- Figure 10 shows that V5 appears to have a slight warm bias over land and cold bias over ice in the daytime. Over Europe, where ECMWF is most accurate, V5 has a warm bias in July and a cold bias in January probably because it is frozen over then. Also, over Antarctica in July the spatial

fluctuations in the temperature bias are larger than in January. The tendencies are similar for nighttime observations as well (Figure 11).

Error Estimates

The χ^2 of the surface temperatures shown in Figure 6 show that the surface temperature error estimates in V5 are significantly better than those in V4. Also, the χ distributions in Figure 12 indicate that the outliers seen in Figure 9 have correspondingly larger error estimates.

Comparison with Dome C

Since ECMWF is presumed to be less accurate over frozen surfaces we have compared a small subset of the AIRS retrieved surface air temperature measurements to in-situ measurements made at Dome C in Antarctica. V5.0.11.0 was used for these comparisons but there is little difference between V5.0.14.0 and V5.0.11.0 at the poles. Figure 13 shows the retrieved surface air temperature (TSurfAir) compared with the 3 meter high air temperature measured at Dome C in Antarctica for observations within 20 km and 15 minutes of a Focus Day over a ~9 month period. The retrieved temperatures in V4 and V5 are similar, however, the V5 temperatures have more realistic error estimates and fewer (actually none) rejected retrievals.

Figure 14 shows surface skin temperature and the surface air temperature for 2 days: one in austral fall, and one in austral winter. The surface temperatures have a slight negative bias while the surface air temperatures have a slight positive bias. These biases are consistent with the boundary layer near Dome C that is dominated most of the year by an inversion with a median 3 meter high air temperature ~1 K warmer than the surface air temperature (Hudson & Brandt 2005). Hudson & Brandt also show that the inversion can be disrupted by surface heating from the sun or downwelling long wave radiation from clouds. This can explain why the retrieved surface temperature matches the three meter high air temperature at the peak temperature of the day on February 9, 2005 (Figure 14) but subsequently is colder.

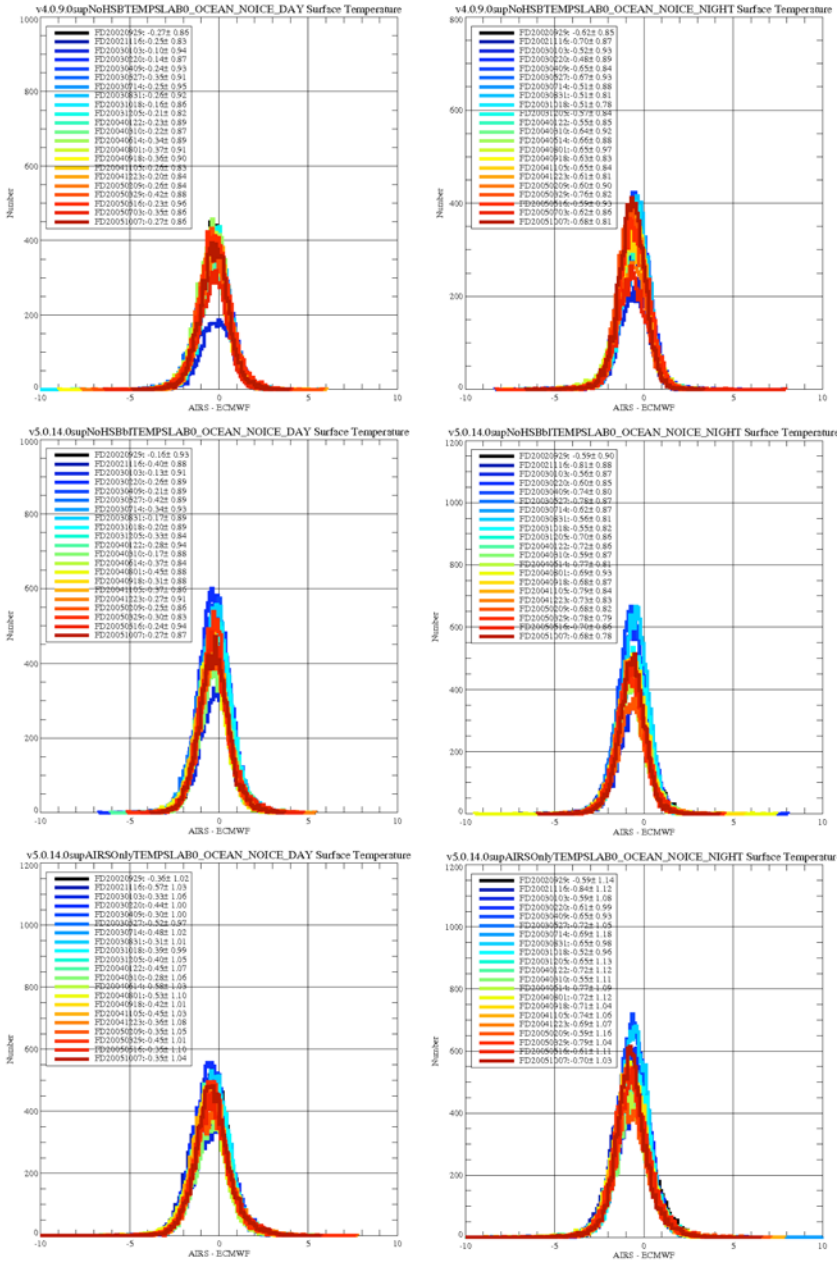

Figure 8 Histograms of the difference between the AIRS and ECMWF Sea Surface Temperatures are displayed with a 0.1 K bin size. The figures on the left are for daytime views the figures on the right are for nighttime views. The top panels show V4, the middle panels show V5 AIRS/AMSU, and the bottom panels show V5 AIRS-Only.

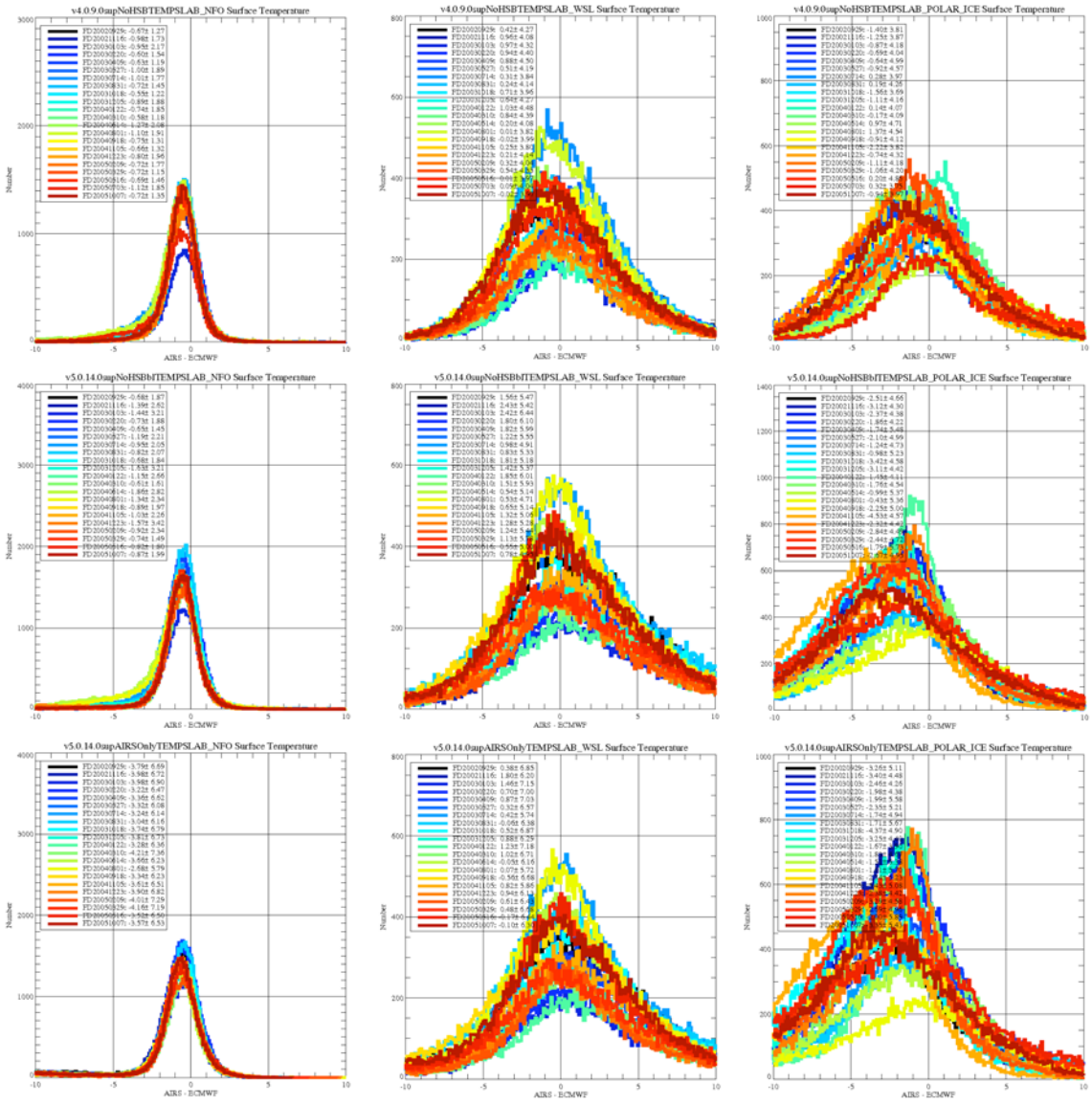

Figure 9 Histograms of the difference between the AIRS and ECMWF Surface Temperature for Non-Frozen Ocean, Warm Season Land, and Polar Ice (from left to right) are displayed. Frozen land at lower latitudes is not displayed because there are large seasonal variations in the sample size. The top panels show V4, the middle panels show V5 AIRS/AMSU, and the bottom panels show V5 AIRS-Only.

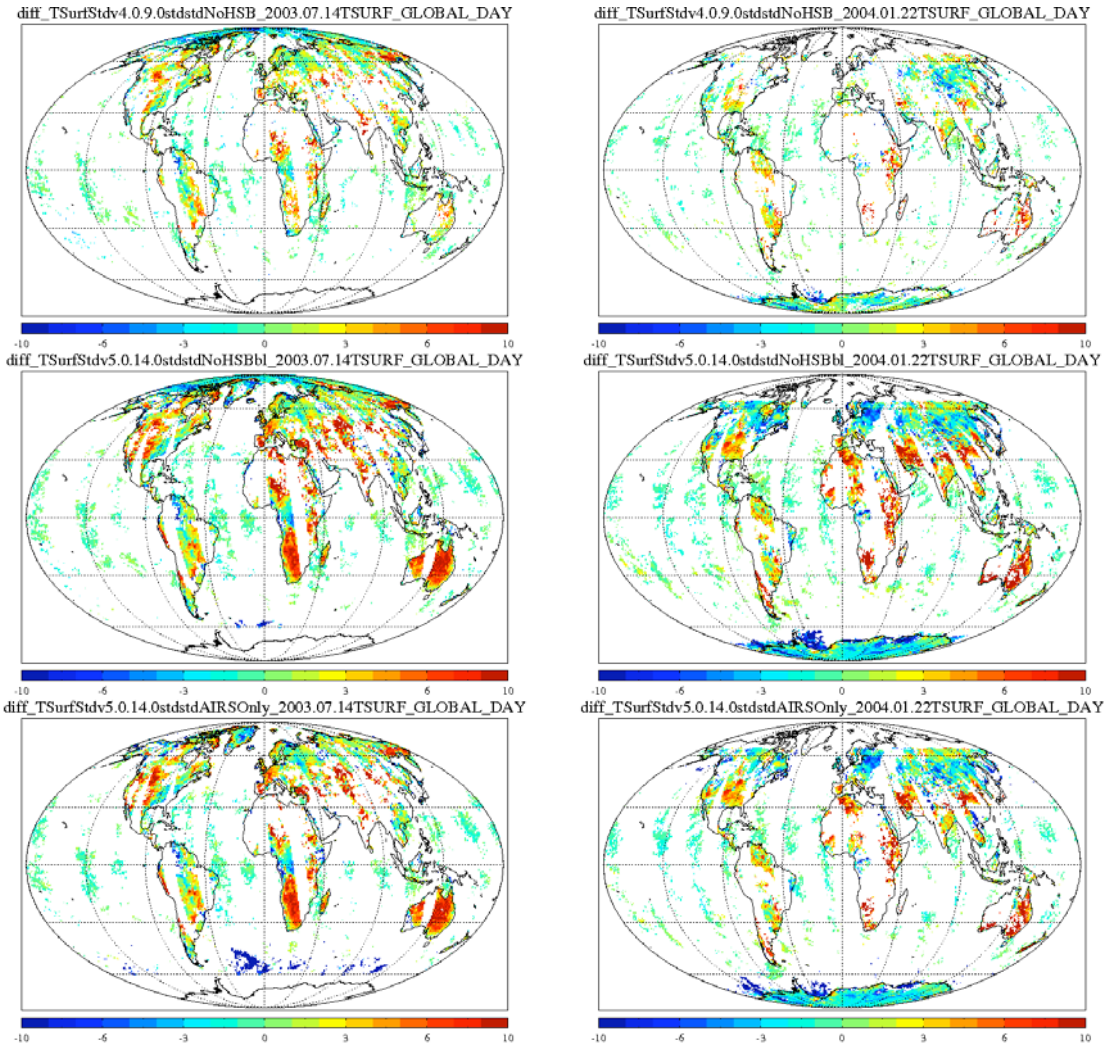

Figure 10 A global map of the difference between that AIRS and ECMWF daytime surface temperatures are displayed for 2 focus days (July 14, 2003 and January 22, 2004).

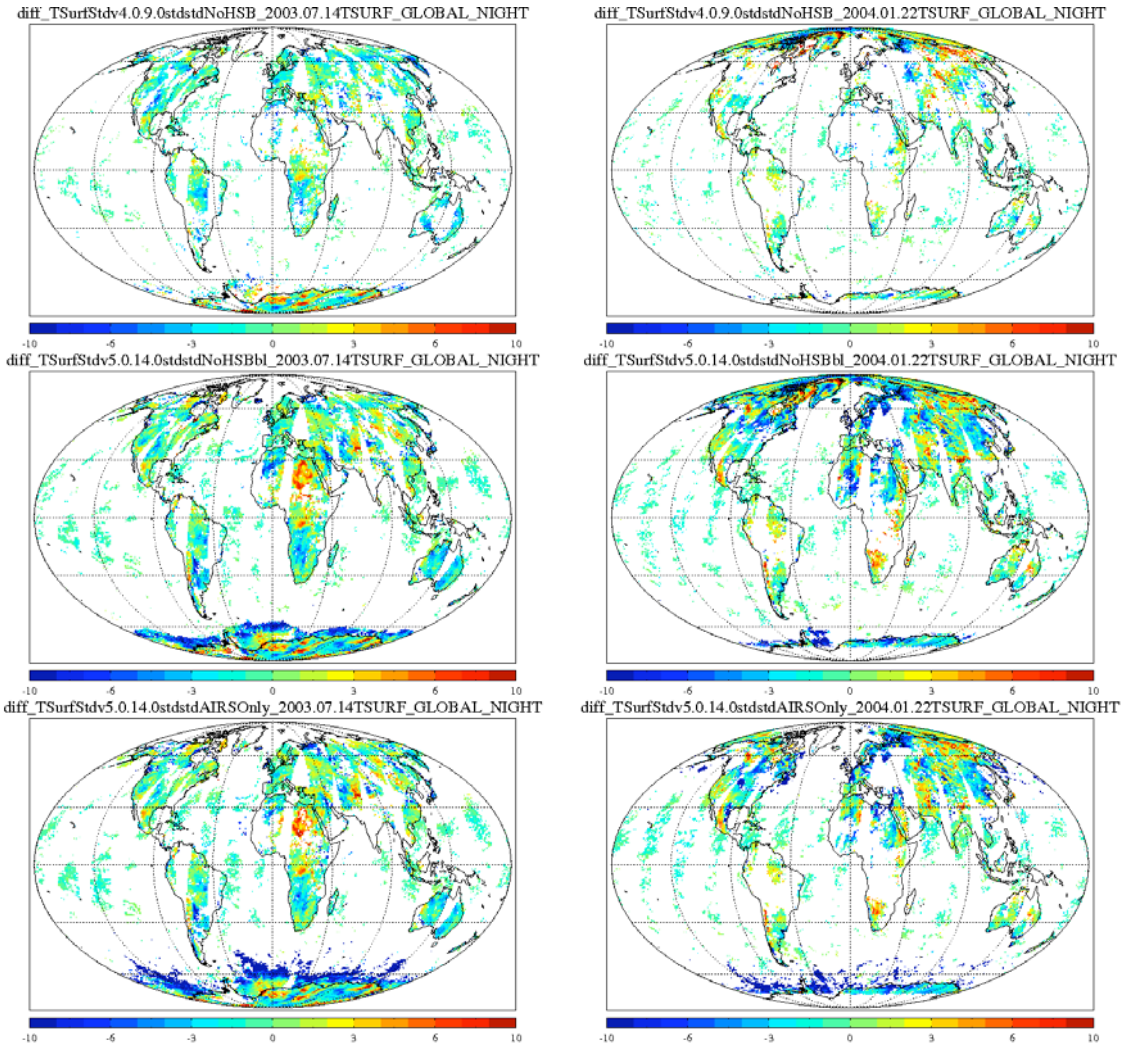

Figure 11 A global map of the difference between that AIRS and ECMWF nighttime surface temperatures are displayed for 2 focus days (July 14, 2003 and January 22, 2004).

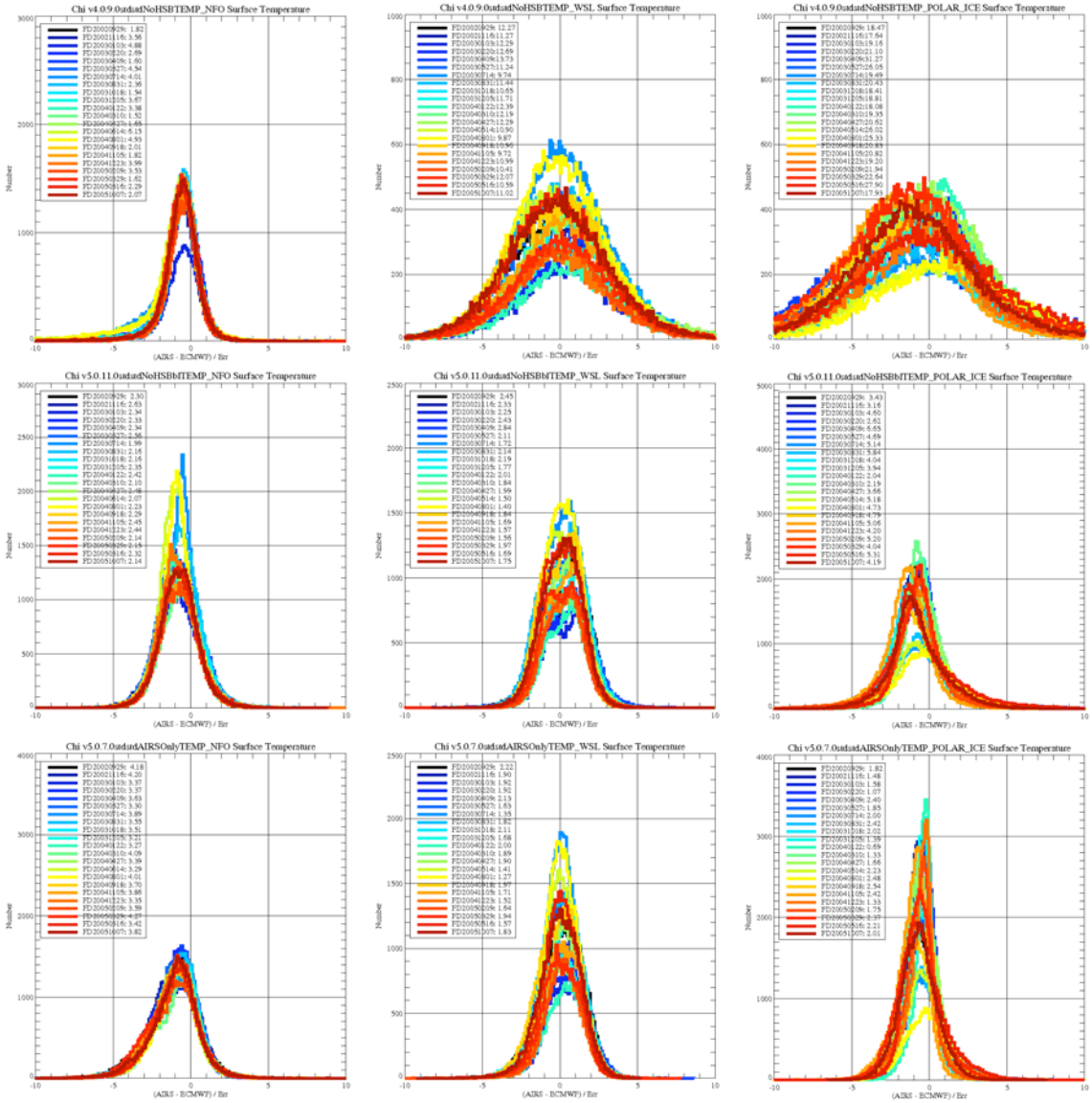

Figure 12 The distribution of χ as defined in Equation (1) is shown for V4 and V5.

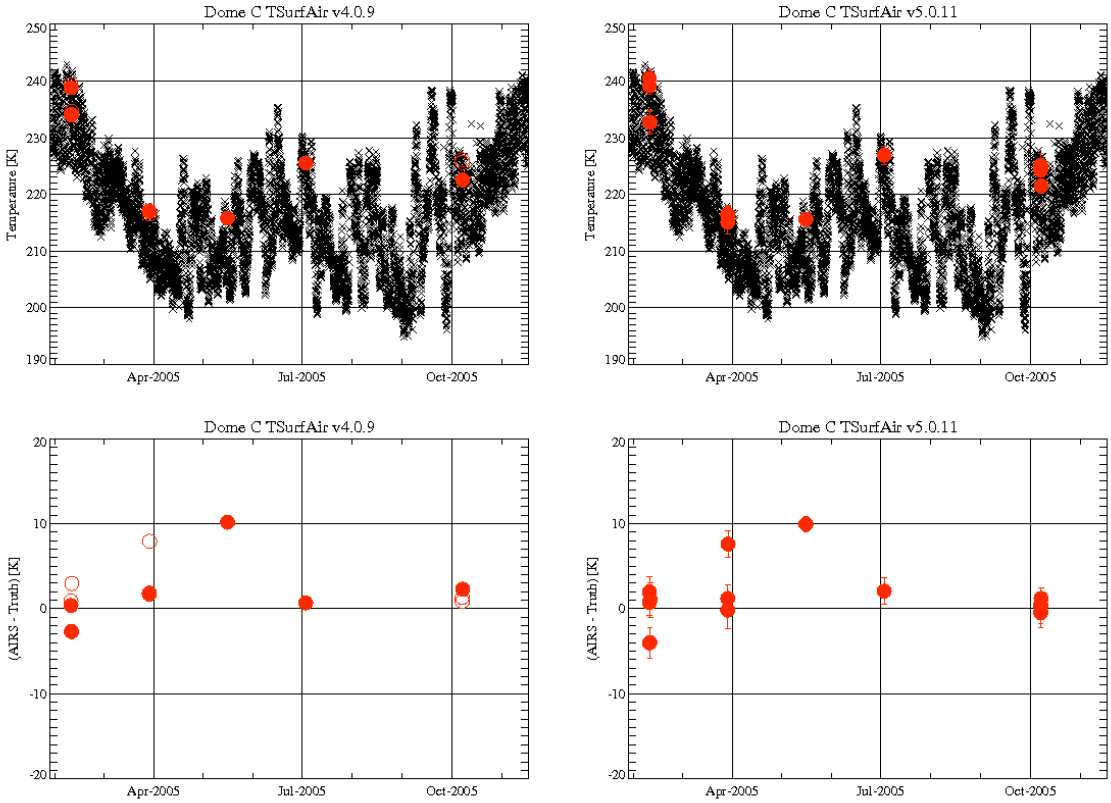

Figure 13 The top panels show a time series of the 3 meter high air temperature measured at Dome C in Antarctica (black) with the AIRS surface air temperature in red. The bottom panels show the residuals between the AIRS measurements and the Dome C measurements. The figure shows TSurfAir from V4.0.9.0 and V5.0.9.0 and the surface temperature measured at Dome C. Open circles are cases with Qual_Surf = 2 and filled circles have Qual_Surf = 0 or 1.

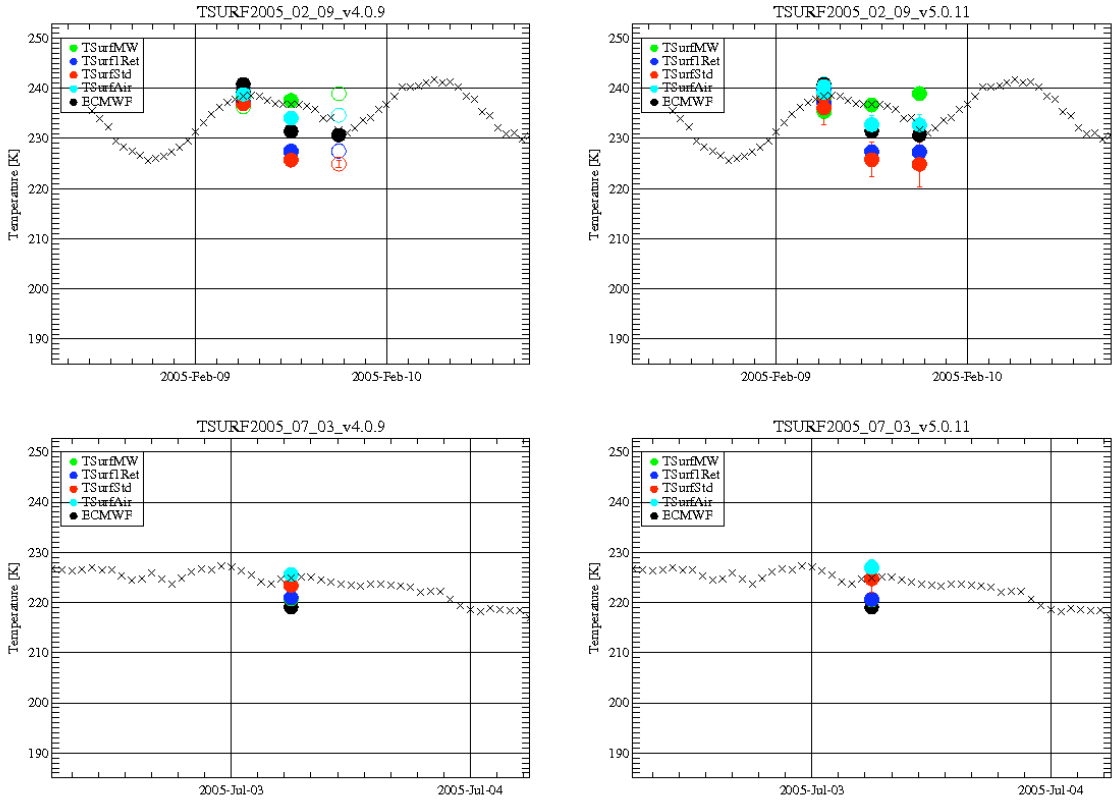

Figure 14 A time series of hourly measurements of the 2 meter high air temperature at Dome C compared with several AIRS temperature retrievals.

4.4. Microwave Surface Products

Parameter List

- **sfcTbMWStd**
- **EmisMWStd**
- **MWSurfClass**
- **EmisMWStdErr**

The Microwave (MW) Only algorithm produces spectral surface brightness and spectral surface emissivities. Although surface skin temperature can be estimated as the quotient of surface brightness (**sfcTbMWStd**) over the emissivity (**EmisMWStd**), it is not a validated product since its accuracy is questionable. There was only a minor improvement in the surface emissivity model since V4. Therefore no significant change is expected in V5.

The V4 MW surface model misidentified precipitation over open water as icy surfaces. This algorithm was modified to correct many of the misclassifications.

The surface skin temperature for any of the 7 spectral frequencies, can be mapped instead of the surface brightness. Some of the issues with the final skin temperature could be caused by biased MW-only skin temperature. So the bias of MW skin temperature over ocean with respect to forecast model will be studied. The complexity of MW spectral surface emissivities is beyond the scope of testing and hence will be postponed until the validation period.

4.5. Water Vapor

Parameter List

- **H2OMMRSat**
- **Qual_H2O**
- **H2OMMRStd**
- **H2OMMRStdErr**
- **totH2OMWOnlyStd**
- **totH2OStd**
- **totH2OStdErr**

The ATLP requires a 15% RMS uncertainty for 2 km thick layers of water vapor in the troposphere and a 5 % RMS uncertainty for the total precipitable water. Although ECMWF water vapor profiles are not of sufficient quality to verify that the AIRS data are meeting these requirements, we have used comparisons with ECMWF to characterize differences with V4.

V5 has fewer anomalously high total water vapor retrievals over warm scenes than there were in V4. However, V5 has more failed retrievals that are filled with –9999 (Figure 15).

Figure 16 and Figure 17 show that the number of water vapor retrievals with Qual_H2O = 0 or Qual_H2O = 0 or 1 is lower in V5 than it was in V4, however, the quality of the retrievals has improved. Radiosonde comparisons have shown that the V4 algorithm produced a dry bias in the upper troposphere of a few percent. Therefore, although Figure 16 and Figure 17 show that V4 matches ECMWF better, the ECMWF water vapor profiles probably have a dry bias. Since the figures show the upper troposphere is wetter in V5 with respect to ECMWF, the dry bias may have improved.

Figure 18 and Figure 19 show maps of the percent difference of water vapor with respect to ECMWF. V5 tends to be drier in the lower troposphere, particularly in the latitude range: –30 to –60.

Figure 20 shows that the V5 water vapor error estimates have improved but that they can still have large chi squares. These could be due to bad retrievals, bad ECMWF data, or underestimates of the error. This should be further investigated in the validation campaign. Figure 21 shows that over land and warm ocean V5 has slightly better retrievals or error estimates but closer to the poles there may be some degradation.

Figure 22 shows that the frequency of outliers is similar in V4 and V5.

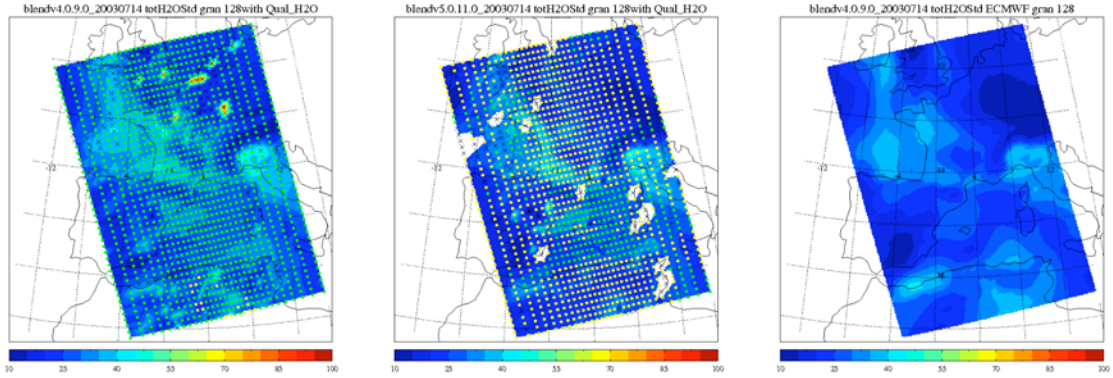

Figure 15 This figure shows that the anomalously high relative humidities seen in V4 (left panel) are no longer found in V5 (center panel). However, V5 has more frequent cases where the retrieved quantity is not in the Standard product. The right panel shows the total water vapor from ECMWF. Green dots are where, Qual_H2O = 0, Yellow Dots are where it is 1, and black X's are where it is 2.

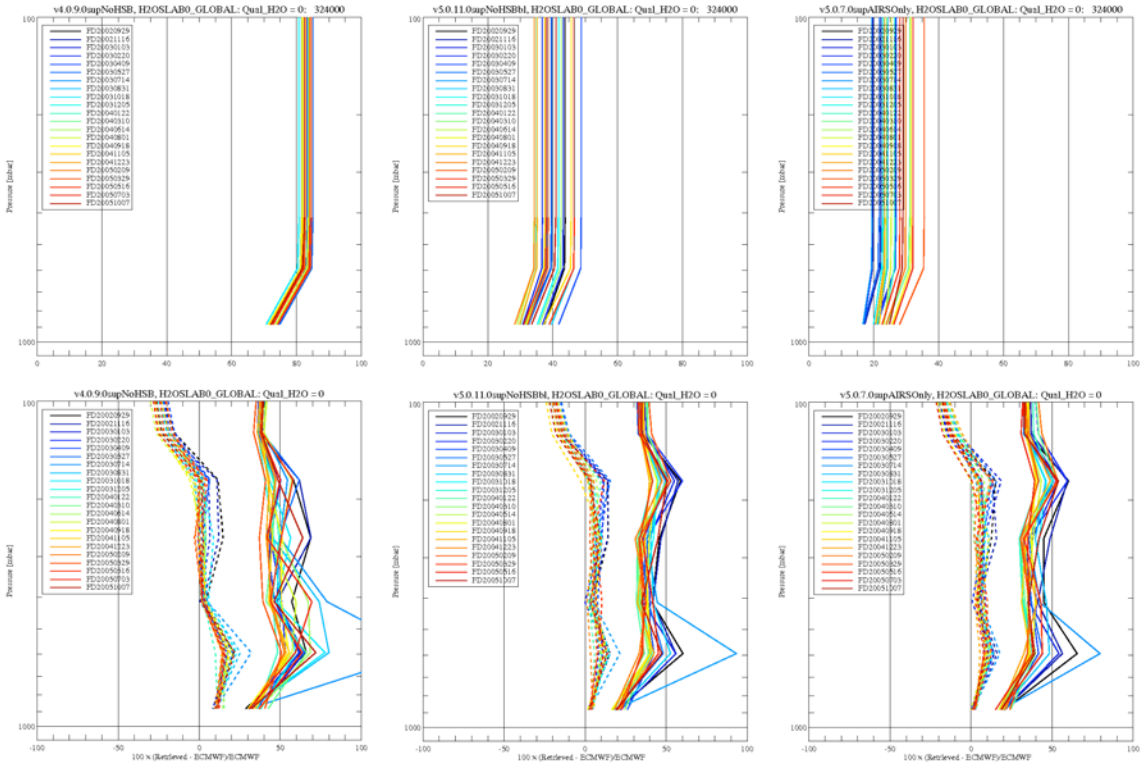

Figure 16 This figure shows the yield, bias and RMS with respect to ECMWF for V4 and V5 water vapor profiles with Qual_H2O = 0.

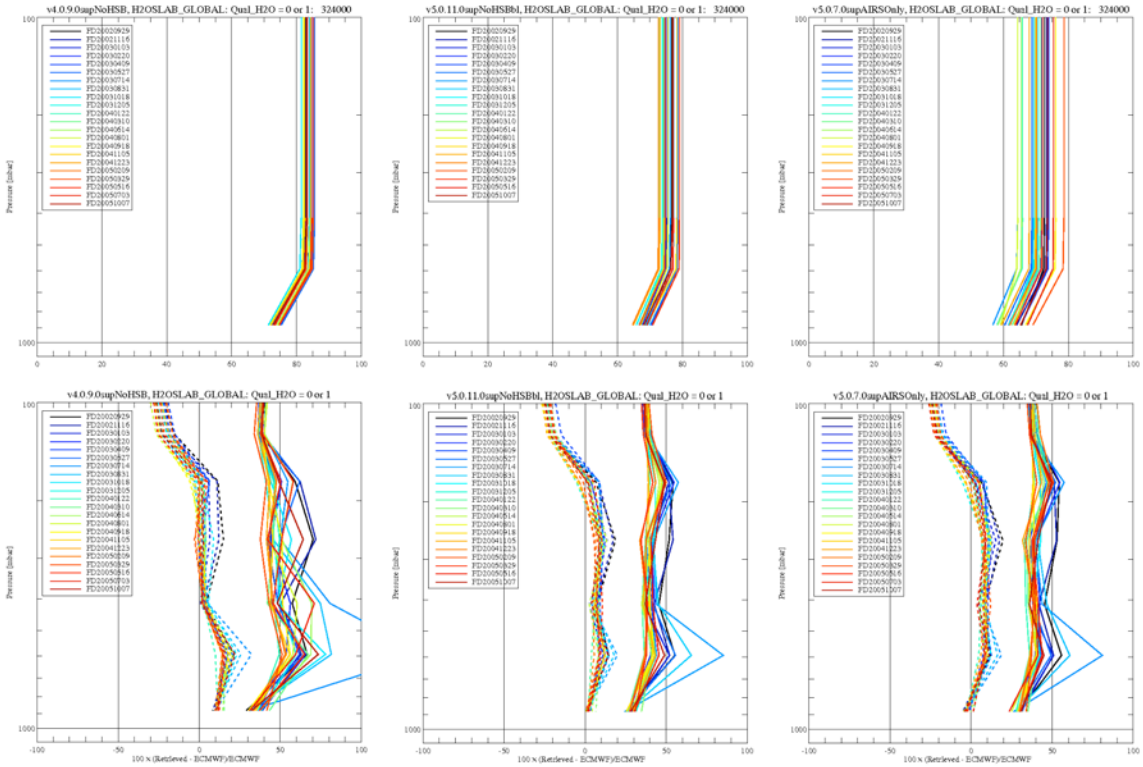

Figure 17 The figure shows the yield, bias, and RMS with respect to ECMWF for the AIRS V4 and 5 retrievals with Qual_H2O = 0 or 1

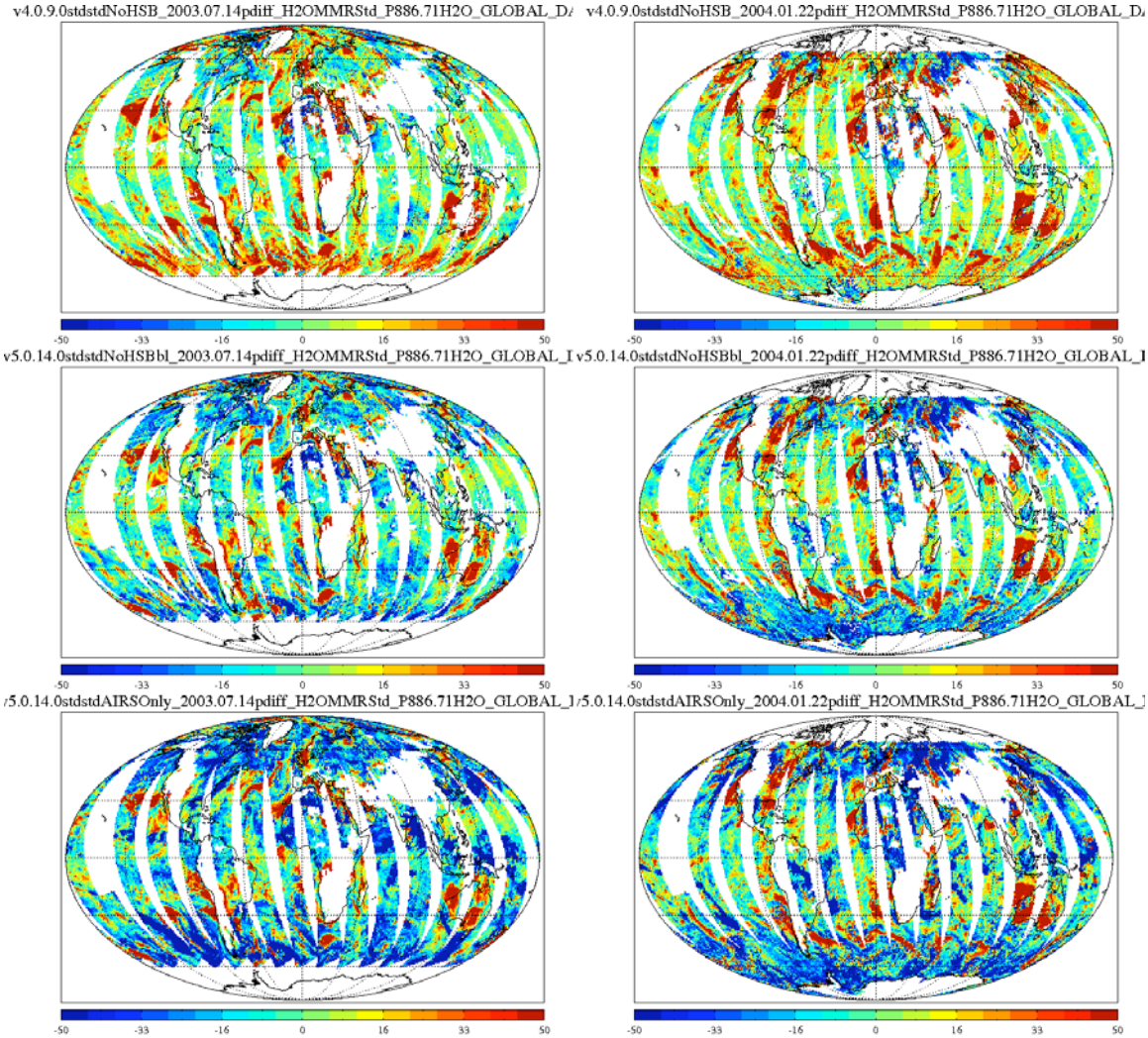

Figure 18 Water Vapor mixing ratio bias (percent difference) in the lower troposphere with respect to ECMWF for the daytime retrievals from 2 days (V4.0.9 [top], V5.0.14.0 AIRS/AMSU [middle], and V5.0.14.0 AIRS-Only [bottom]).

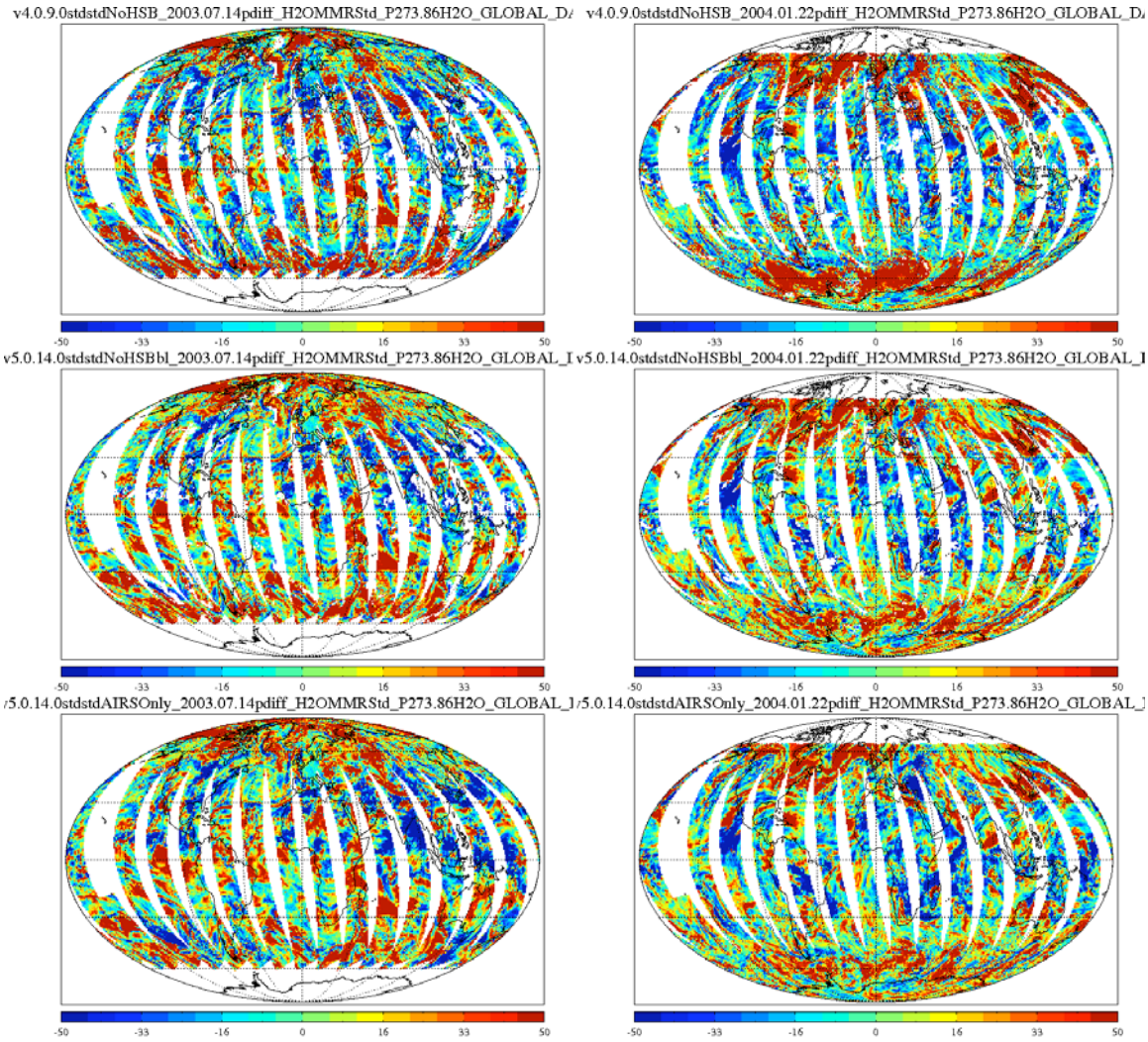

Figure 19 Water Vapor mixing ratio bias (percent difference) in the upper troposphere with respect to ECMWF for the daytime retrievals from 2 days (V4.0.9 [top], V5.0.14.0 AIRS/AMSU [middle], and V5.0.14.0 AIRS-Only [bottom]).

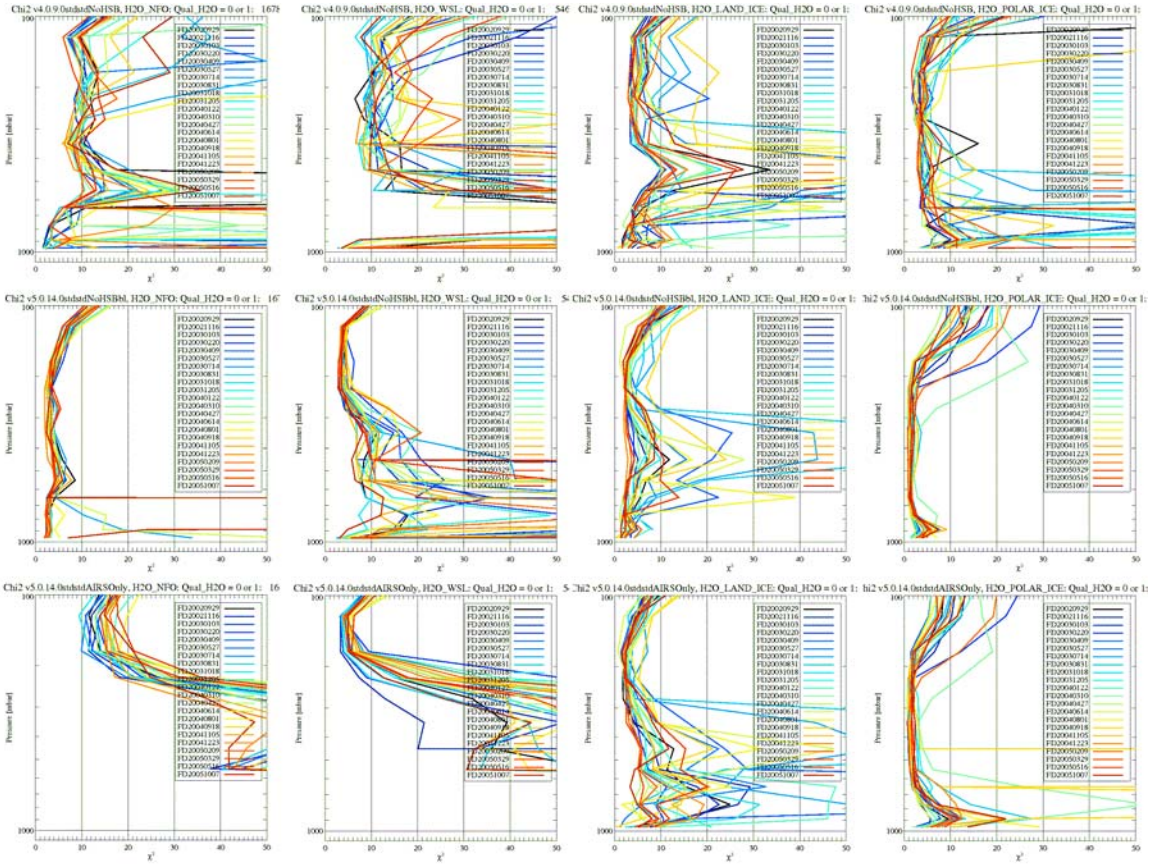

Figure 20 The figure shows the Reduced chi square calculated for 4 climatological regions. The top panels are based on V4.0.9.0, the middle panel is V5.0.14.0 AIRS/AMSU, and the bottom panel is V5.0.14.0 AIRS-Only.

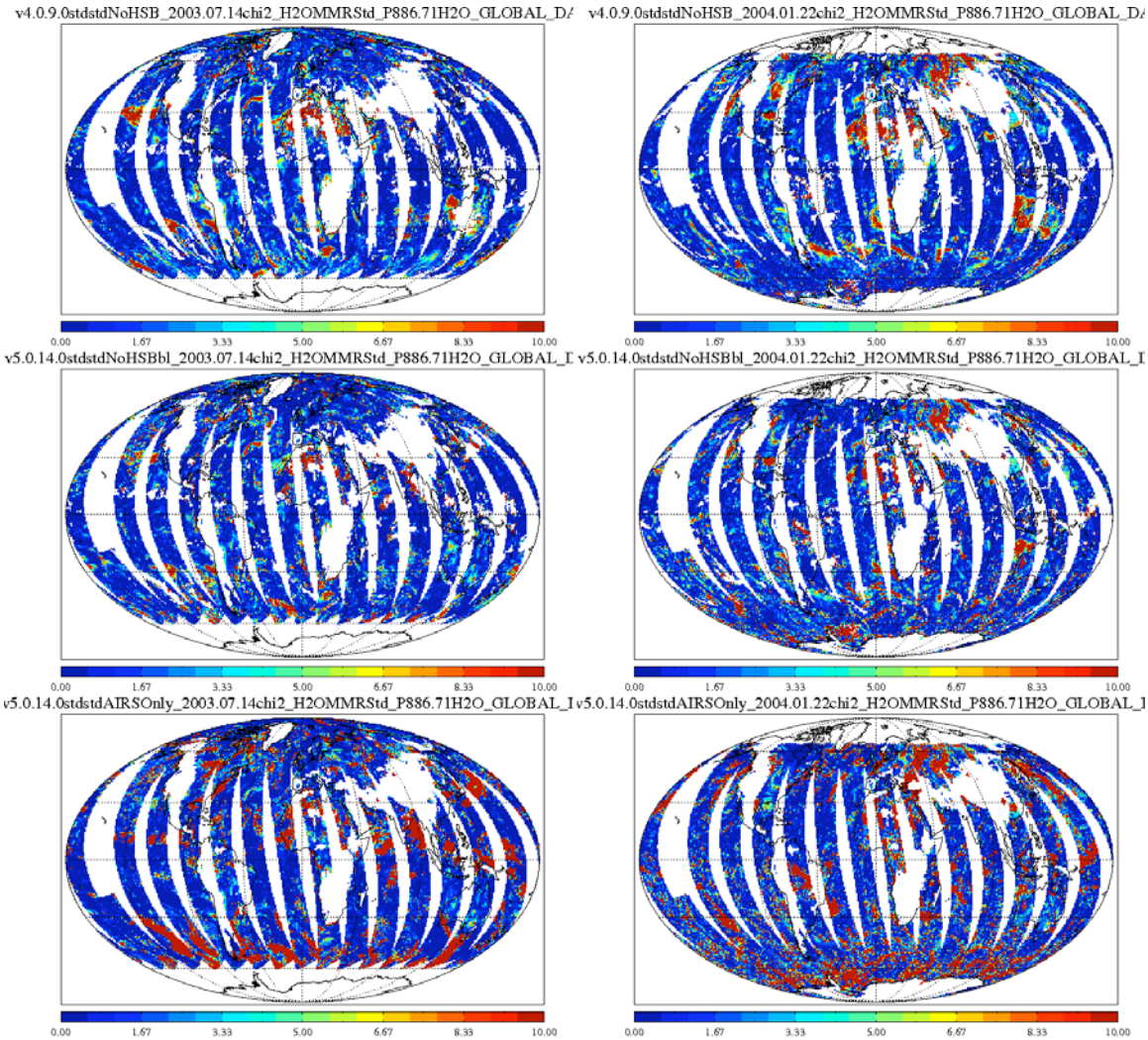

Figure 21 The figures show the spatial distribution of the square of the daytime AIRS-ECMWF retrieved mixing ratio divided by the error estimate in the lower troposphere for 2 days (July 14, 2003 and January 22, 2004) using the V4.0.9.0 (top), V5.0.14.0 AIRS/AMSU (middle), and V5.0.14.0 AIRS-Only (bottom). The V5 AIRS/AMSU error estimates are better than those from V4 but the AIRS-Only error estimates need improvement.

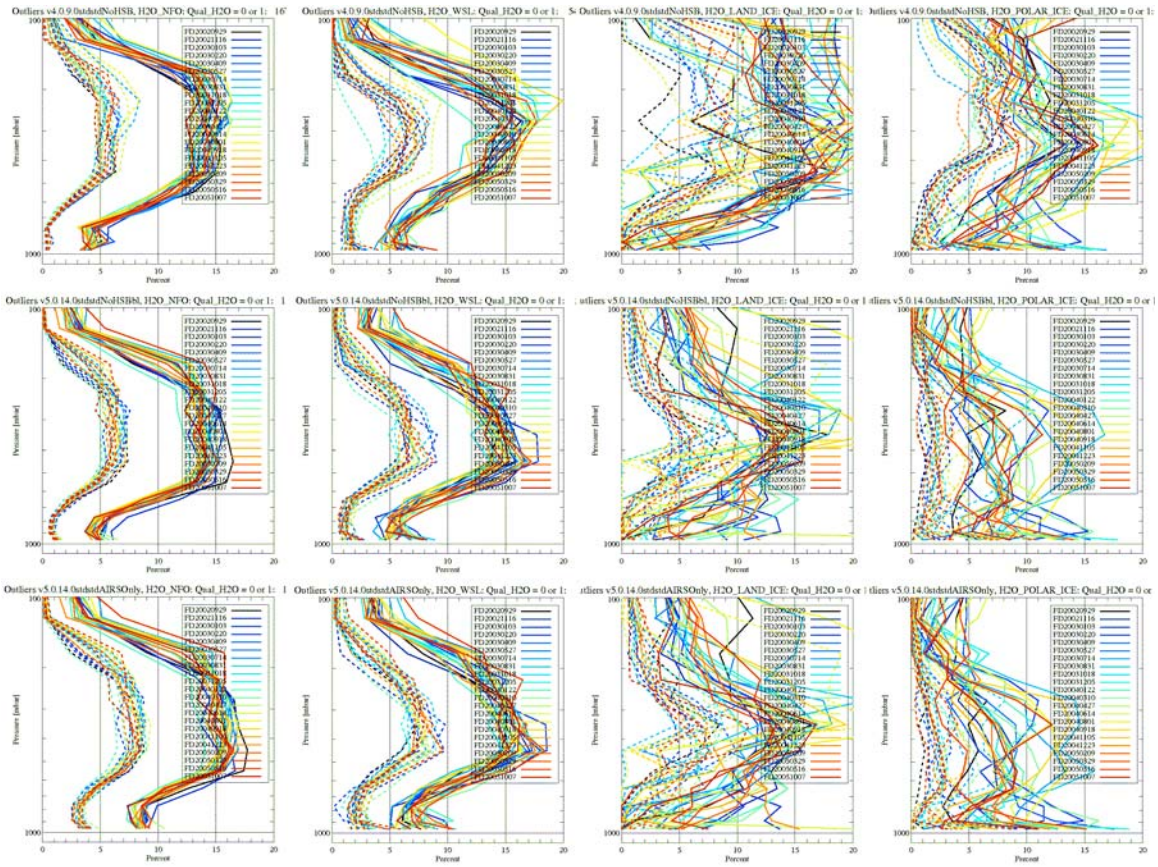

Figure 22 The percentage of inner and outer outliers is displayed for the four climate regimes (NFO, WSL, Land Ice, Polar Ice, left to right) and three versions (V4.0.9.0, V5.0.14.0 AIRS/AMSU, V5.0.14.0 AIRS-Only, top to bottom)

4.6. Cloud Parameters

Parameter List:

- **totCldH2OStd**
- **totCldH2OStdErr**
- **numCloud**
- **TCldTopStd**
- **TCldTopStdErr**
- **PCldTopStd**
- **PCldTopStdErr**
- **CldFrcStd**
- **CldFrcStdErr**

As part of the evaluation of V5 cloud products, we compare the AIRS cloud height and fraction fields to the active cloud measurements made by CloudSat (Stephens et al. 2002). CloudSat is a 94 GHz profiling radar that measures vertical profiles of clouds and precipitation with an approximate vertical resolution of 250 m, and the horizontal footprint is approximately 1x2 km. Several products are produced operationally, including a cloud mask with confidence intervals, ice and liquid water content, latent heating fluxes, and a cloud classification scheme, among others. AIRS and CloudSat are offset by about 55 seconds; this is an especially important feature in eliminating cloud evolution effects as a source of cloud field discrepancies common in comparisons of satellite and surface-based observations (Kahn et al. 2005). CloudSat does have some observing limitations beyond its limited nadir sampling. (1) The lowest 1 km or so of observations is hindered by surface effects that make the cloud retrieval unreliable. (2) The 94 GHz frequency is not sensitive to small hydrometeors common in thinner cirrus clouds near the tropopause and non-precipitating stratus clouds, and these are commonly missed by CloudSat. (3) Even though CloudSat can profile through most clouds, the signal is attenuated by precipitation. However, this is expected to have very little to no impact in AIRS infrared-based retrievals because the IR signal saturates more quickly than cloud radar.

In this evaluation we confine ourselves to comparisons of the CloudSat cloud mask to the AIRS 2-layer cloud top height, using the CloudSat cloud classification product to partition the cloud fields by type (There are 8 cloud classifications: Cirrus (Ci), Altostratus (As), Altocumulus (Ac), Stratus (St), Stratocumulus (Sc), Cumulus (Cu), Nimbostratus (Ns), and Cumulonimbus (Cb)). The AIRS cloud height is calculated from the cloud top pressure fields using adjacent (in the case of opaque clouds) or coincident (in the case of tenuous clouds) temperature profiles to calculate height. Figure 23 shows an illustrative example of the AIRS 2-layer cloud height retrieval superimposed on the

CloudSat cloud mask for AIRS V4 cloud retrievals. The cloud mask is associated with confidence intervals; it is generally thought that values of 20 or more are almost certainly cloud, and values of 10 or more are likely to be cloud. Many values from 6–10 are at the edge of cloud sensitivity, or are false signals due to ground clutter or other effects. Qualitatively, AIRS seems to place clouds in generally the right places. Although where multi-layer cloud structure is noted, the agreement is poorer. Figure 16 shows the same scene except the cloud retrievals are for V5. Note the slight changes in placement of clouds, especially for clouds with low values of Effective Cloud Fraction (ECF) near the surface and in the upper troposphere associated with thinner cirrus clouds. In Figure 24 we show for cloud mask values above 10 the cloud classification of cloud type. This product is combined from MODIS radiances, ECMWF model profiles, and CloudSat reflectance and cloud mask profiles. This figure shows a tendency for better agreement for some cloud type over others. This is a motivating factor to consider a much larger set of data to look at global-scale differences as a function of cloud type.

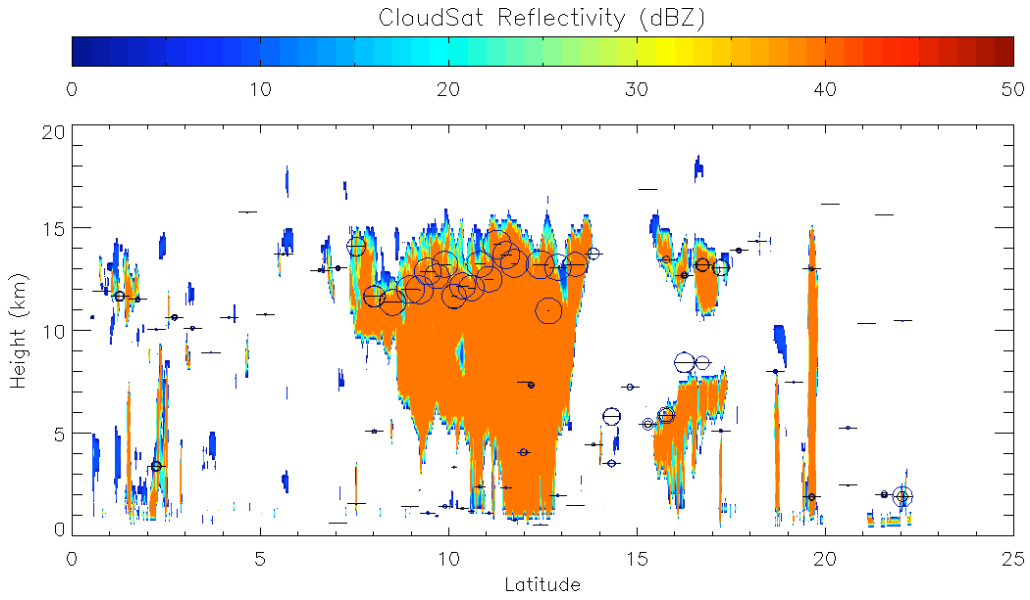

Figure 23 CloudSat cloud mask with confidence values colorized for a granule in the tropical Western Pacific. V4.0.9 AIRS 2-layer height retrieval is superimposed as horizontal lines. The Effective Cloud Fraction (ECF) is linearly proportional to the diameter of circles, with larger diameters equal to larger ECF. CloudSat mask values of 10 or less are highly questionable because of reduced sensitivity and contributions from noise. The authors of the CloudSat cloud mask tool are Jay Mace and Roger Marchand.

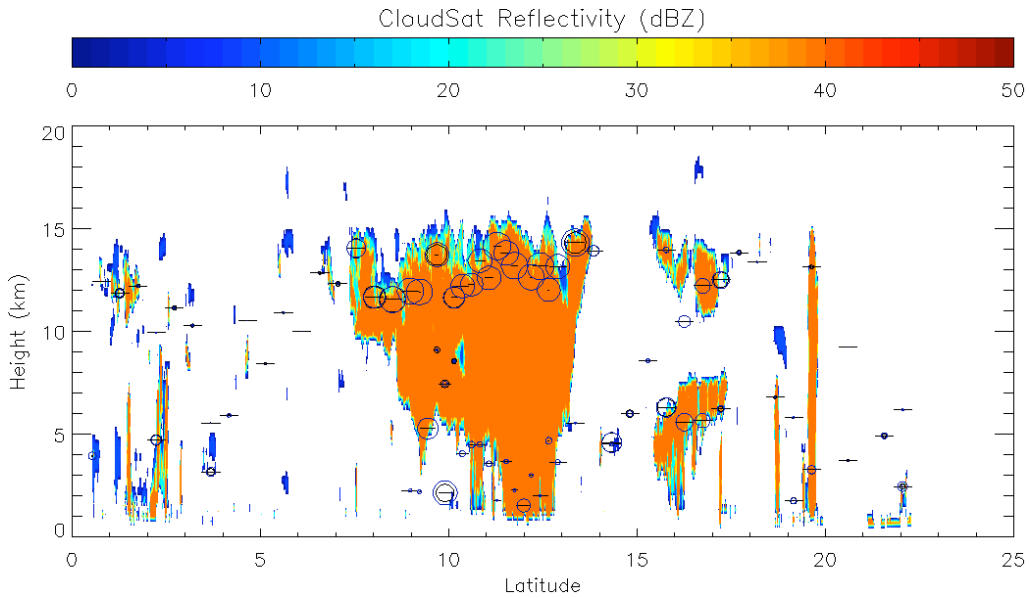

Figure 24 The same as Figure 16 but for V5 retrievals

In Figure 25 we show joint Probability Density Functions (PDFs) of AIRS-CloudSat height differences for three days globally (7-22-2006, 8-15-2006, and 10-26-2006). The height difference takes the highest AIRS cloud layer with the highest CloudSat cloud mask value > 6 and > 10. What is clear is that the peak number of values is centered near AIRS-CloudSat ~0 km, with significant scatter to positive and negative values. When cloud mask values > 10 are used, many of the positive differences disappear because many of the CloudSat-measured clouds at lower values are spurious. The degree of scatter is larger for lower values of ECF, and maximizes below 0.05. A negative tail is observed at low values of ECF, indicating some systematic behaviors likely related to cloud type. Figure 26 shows a more quantitative average and one standard deviation of the variability in the differences. The biases are largest when the ECF values are lowest, and the variability decreases with increasing ECF as well. This is true for both the lower and upper AIRS height differences. This further shows the complexity of cloud types and amounts in interpreting the agreement.

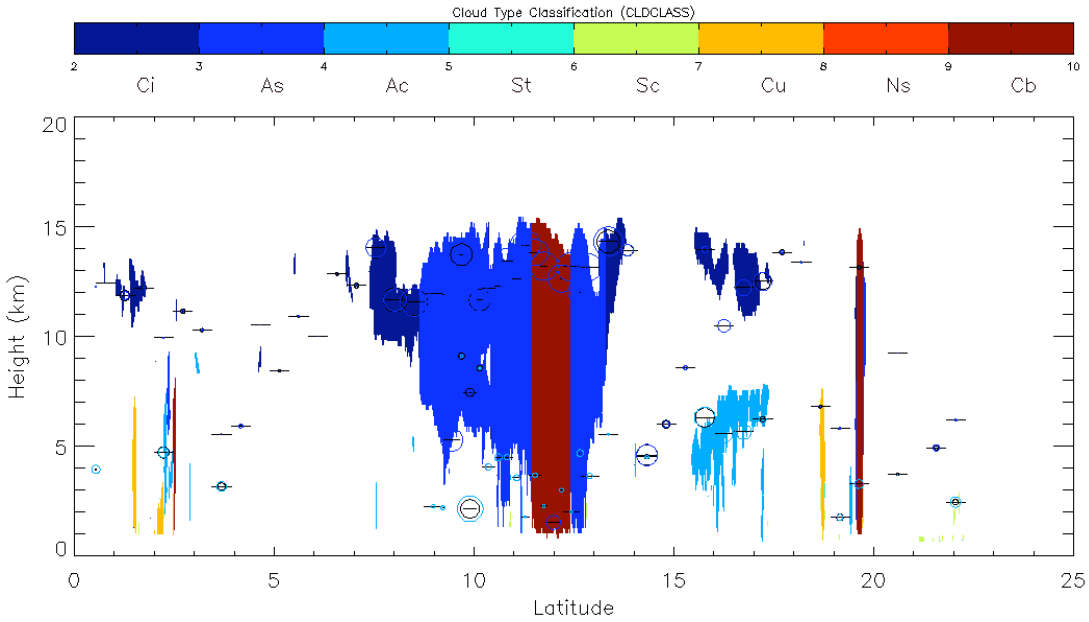

Figure 25 Colorized values represent the cloud types retrieved by the cloud classification product developed by the CloudSat science team. The colorized regions are shown only for cloud mask values of 10 and greater. The 8 cloud classification categories are Cirrus (Ci), Altostratus (As), Altocumulus (Ac), Stratus (St), Stratocumulus (Sc), Cumulus (Cu), Nimbostratus (Ns), and Cumulonimbus (Cb). The authors of the cloud classification algorithm are Zhien Wang and Ken Sassen.

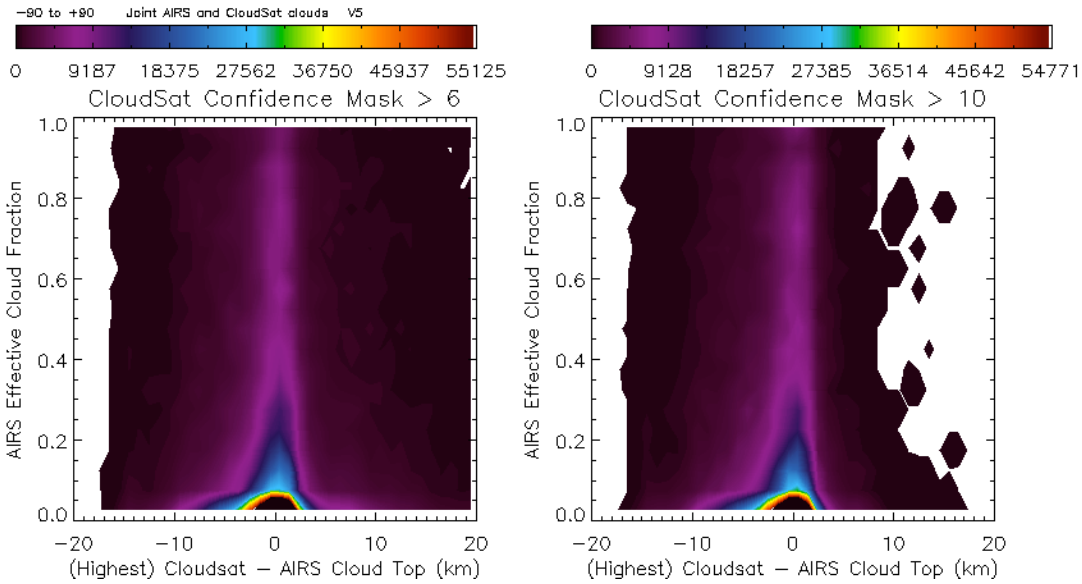

Figure 26 Joint PDFs for CloudSat-AIRS height differences (using the AIRS upper layer). PDFs using Cloudsat cloud mask values >6 and > 10 are shown.

In Figure 27 and Figure 28 we show the individual agreements as a function of the cloud types. Some of the types are differenced with the upper AIRS layer (Ci, Cb, Ns, Ac, As) and the lower AIRS layer (Sc, Cu). Overall, the biases and variability are significantly reduced over all clouds together shown in Figure 26. The results for V4 are shown in all plots as well. As a whole, there is little change (with a slight reduction in bias and variability) compared to V5. However, some of the changes are a bit more substantial for the individual cloud types, including Sc, As, and Ac. Both biases and variability are reduced in V5. Overall, the picture regarding the realism of AIRS clouds is promising. However, there may be a decrease in the sensitivity to low clouds over ocean in some cases (Figure 30).

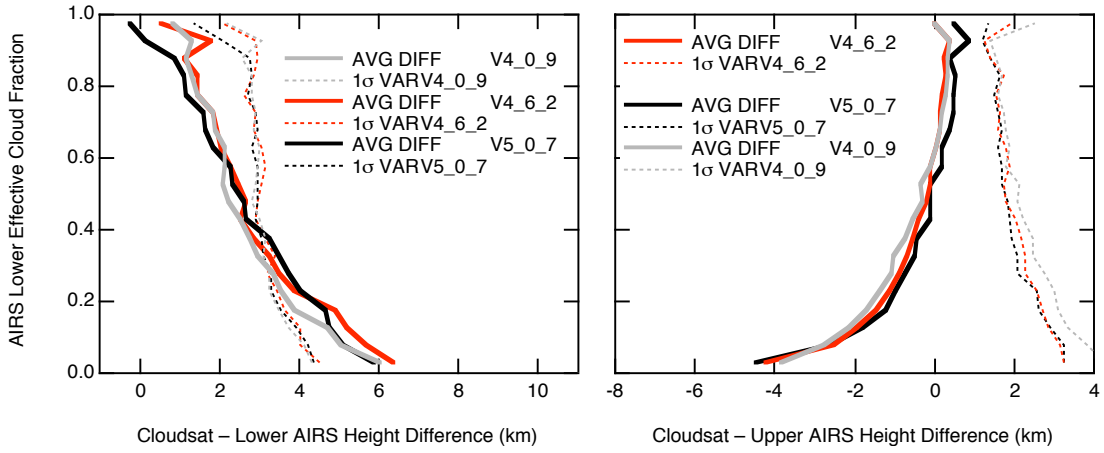

Figure 27 Average and one standard deviation in agreement shown for all clouds. On the left: CloudSat-AIRS difference using the lower layer of airs. On the right: same as the left except the upper layer is used. Both figures are for cloud mask values > 10.

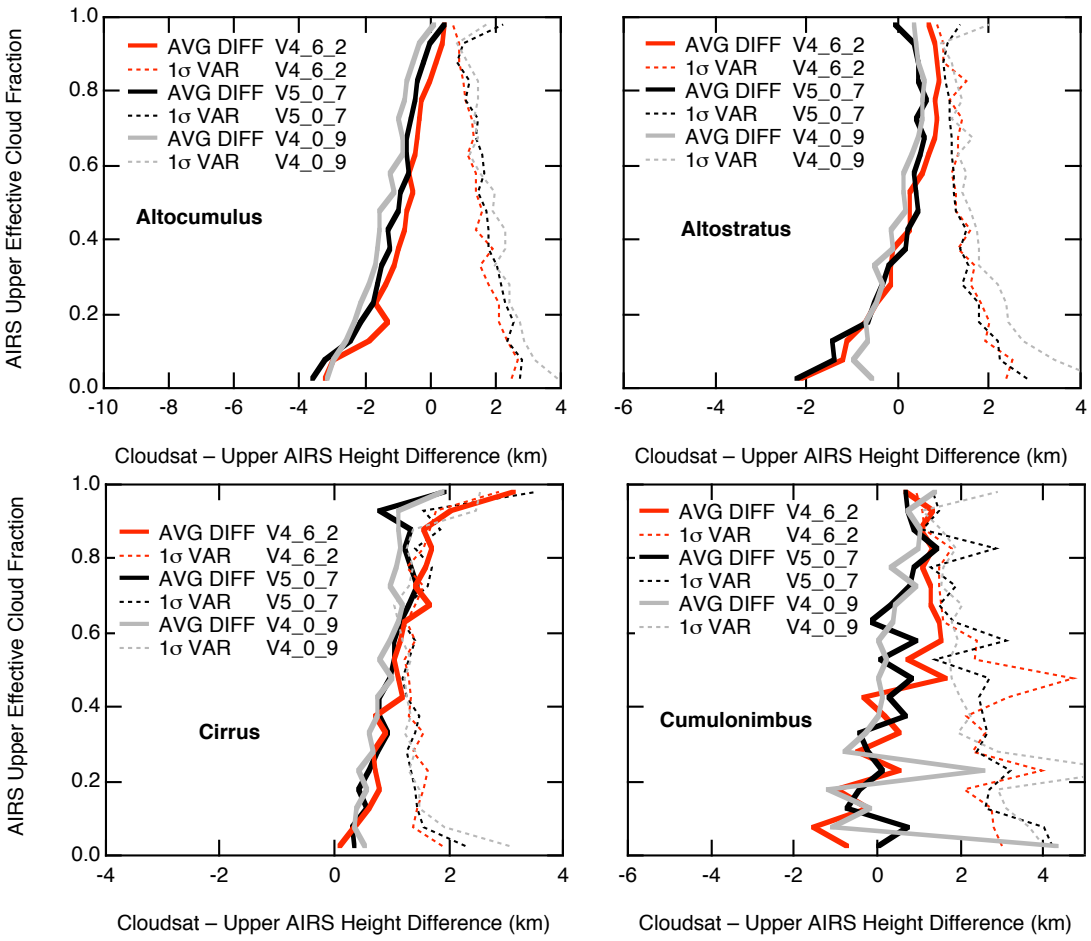

Figure 28 Same as Figure 27 except partitioned by "cloud type". Note that the bias and variability decrease considerably when viewing by cloud types individually. This is largely the case because some lower-layer AIRS clouds agree better with CloudSat, and in other cases, the higher layer agrees better.

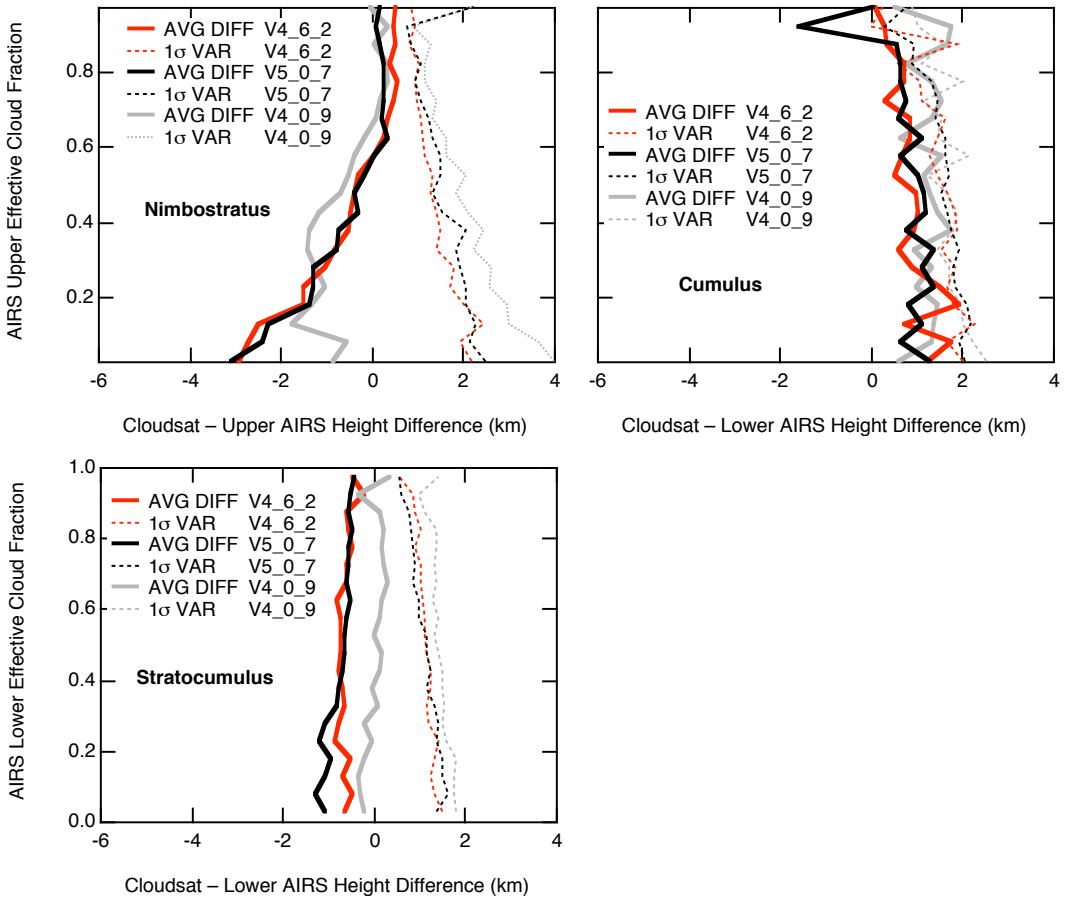

Figure 29 Some additional individual cloud types.

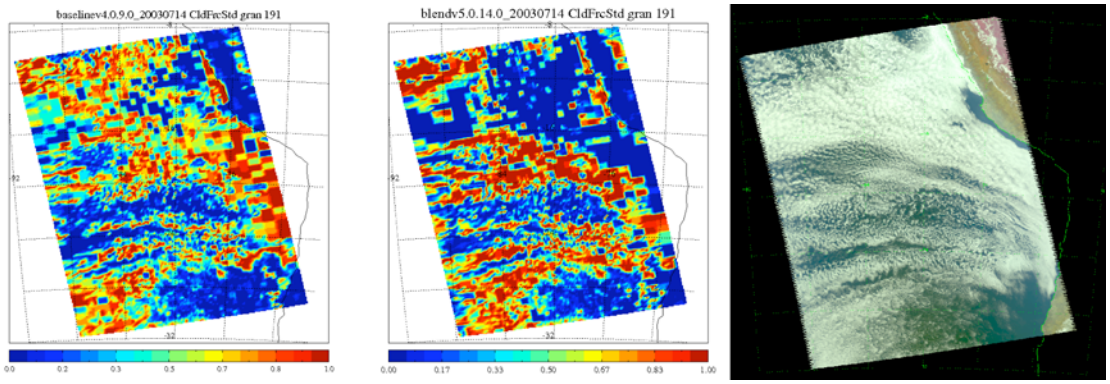

Figure 30 The figures on the left show maps of the retrieved cloud fraction from V4 and V5, respectively. The figure on the right is a visible image of the same region.

4.7. Outgoing Longwave Radiation

Parameter List:

- **Qual_Cloud_OLR**
- **olr**
- **olr_err**
- **clrolr**
- **clrolr_err**

A significant bug was found in V4.0.9 which overestimates cloudy OLR products when there are high clouds in the field of view. The clear sky OLR products were not affected by this bug. This bug was corrected in V4.4.4 (unreleased) and carries through to the V5 release.

The outgoing longwave radiation (OLR) and clear sky OLR are the main products. These parameters are computed from AIRS retrievals using a radiative transfer algorithm. In V3, OLR products were not computed from retrievals of type 100, where the MW Only retrieval fails. This caused bias in OLR field since many such cases have high clouds with low values of OLR. In V4, we attempted to compute OLR products even for the retrievals of type 100, using the failed MW only retrieval. However, a bug in the software caused the OLR algorithm to use cloud height lower than the actual value, producing higher value of OLR. This software bug did not affect the clear sky OLR.

For V4.6.2 and earlier, the error estimate of OLR and clear sky OLR were set to 2 Watts cm^{-2}, without regards to quality of OLR. And there is no immediate plan to improve this crude error estimate.

Figure 31 shows maps of OLR for Sept 6, 2002 from V4.0.9. The maps were generated from level 3 files. The bug mentioned above caused the OLR fields to be too high at certain locations, especially near the Intertropical Convergence Zone (ITCZ). The figure also shows, this bias was mostly removed in V5.0.7, as expected. Also as expected, the comparison of Figures 24 and 25 shows only minor differences in clear sky OLR. The map of the OLR from Aqua CERES shows that AIRS and CERES products are quite similar. The major difference is over land, where AIRS retrieves clear sky OLR while CERES does not (Figure 32).

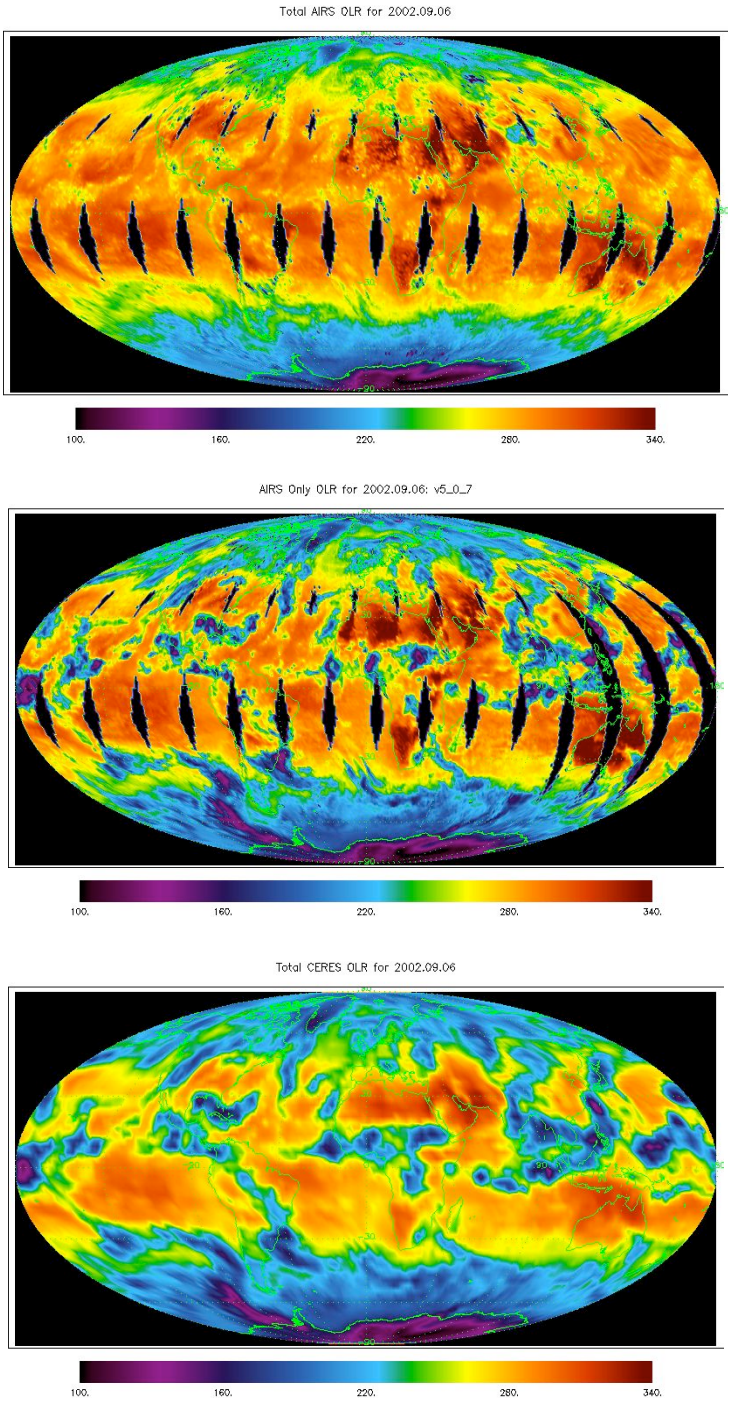

Figure 31 Total OLR from V4 (top), V5 (middle), and CERES (bottom).

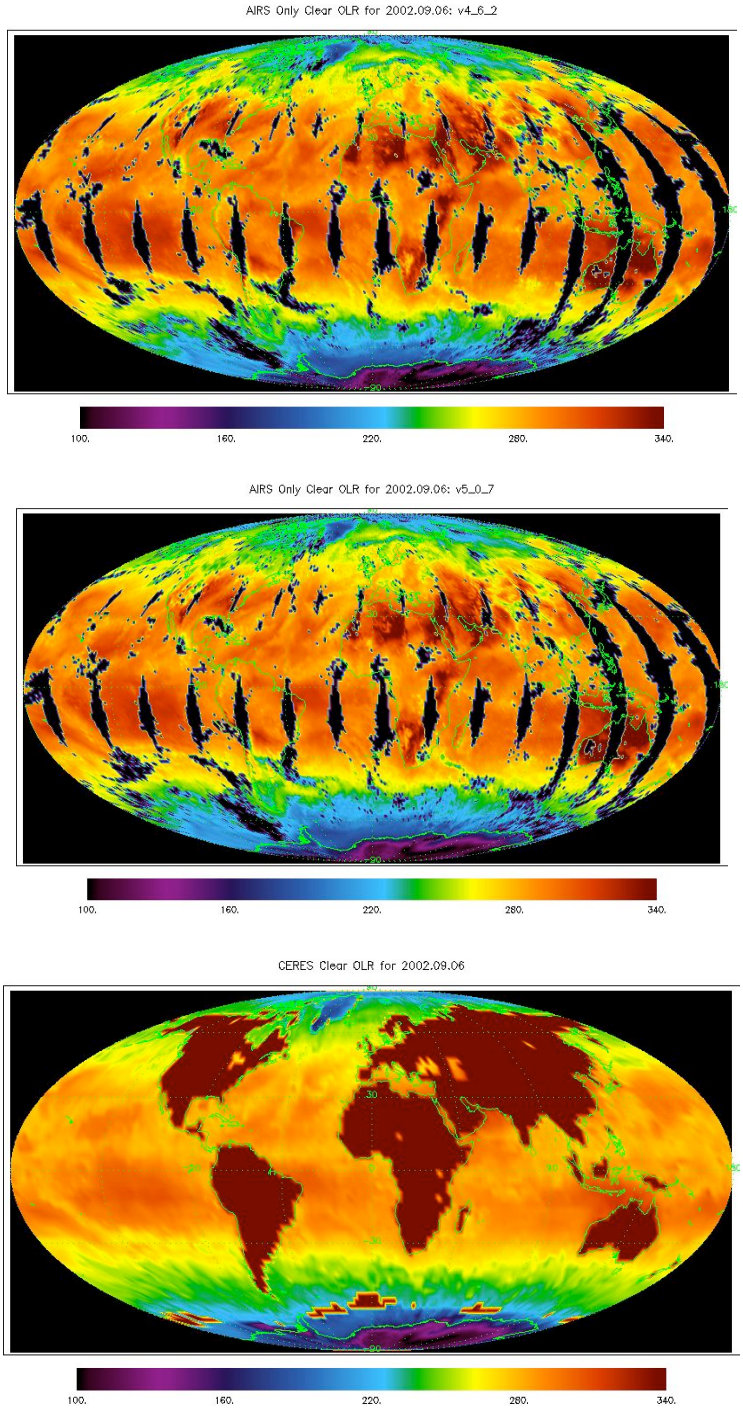

Figure 32 Clear Sky OLR from V4 (top), V5 (middle), CERES (bottom).

The zonal means of AIRS OLR are plotted in Figure 33. The figure shows that the large positive bias in AIRS V4.0.9 total OLR is removed in V4.6.2 and that the V4.6.2 AIRS OLR still has small positive bias of about 11 Watts/meter^2, but comparable, to CERES products.

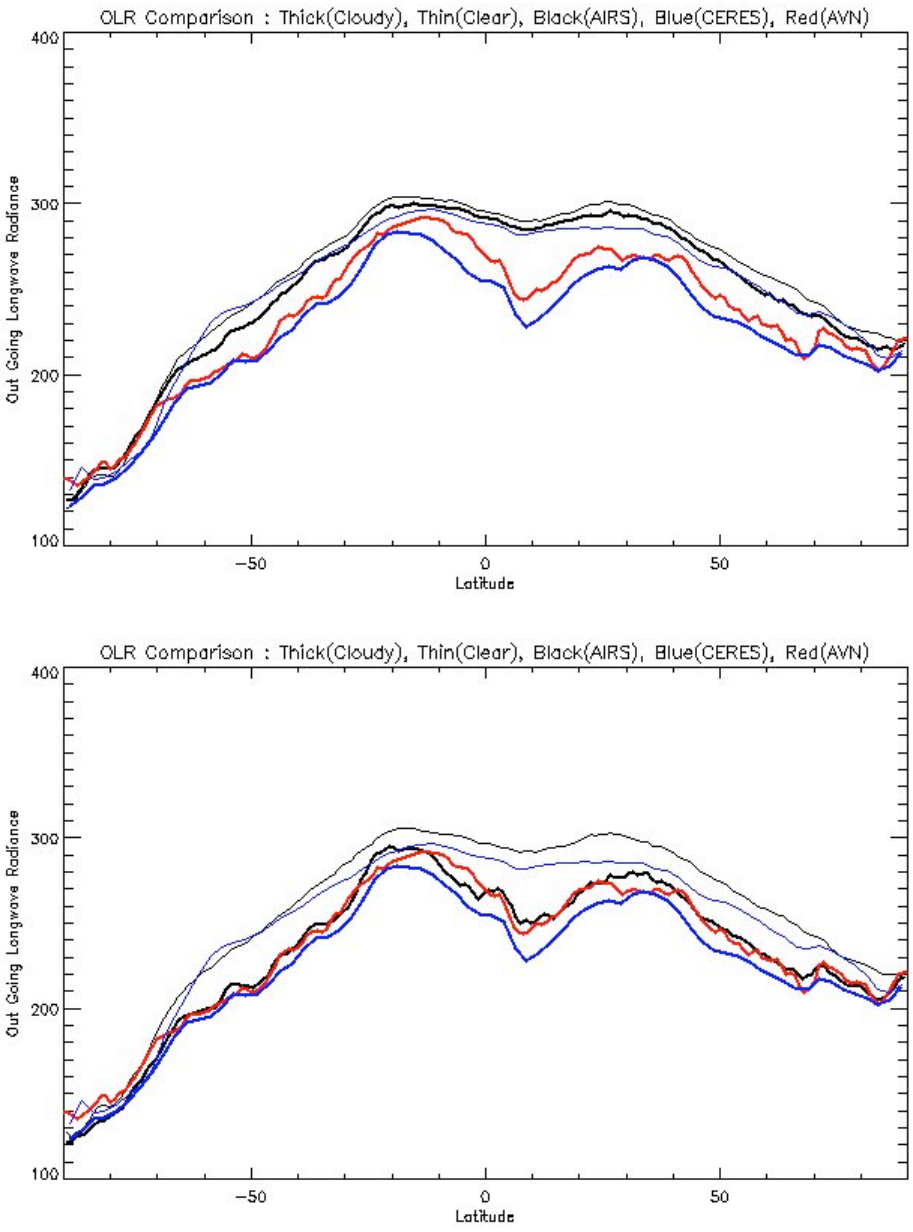

Figure 33 Zonal Mean Plots for AIRS V4.0.9 (top) and V5.0.7 (bottom).

Figure 34 shows the trend of global mean OLR for 48-day-cycle focus days. The global mean was computed from V5.0.7 level 3 products. The black curve represents global mean OLR for AIRS, average of ascending and descending means. The red curve is NCEP's AVN OLR, which was calculated from a 3 hour forecast for each of the focus days. The green curve is for CERES OLR. There is a very good agreement between AIRS and AVN. However, AIRS has an almost constant 10 watts per meter squared bias with respect to CERES. The bias may be attributed to the outdated forward model AIRS uses as well as slight difference in definition of OLR. AIRS is the average of 1:30 AM and 1:30 PM OLR, while CERES values represent daily mean.

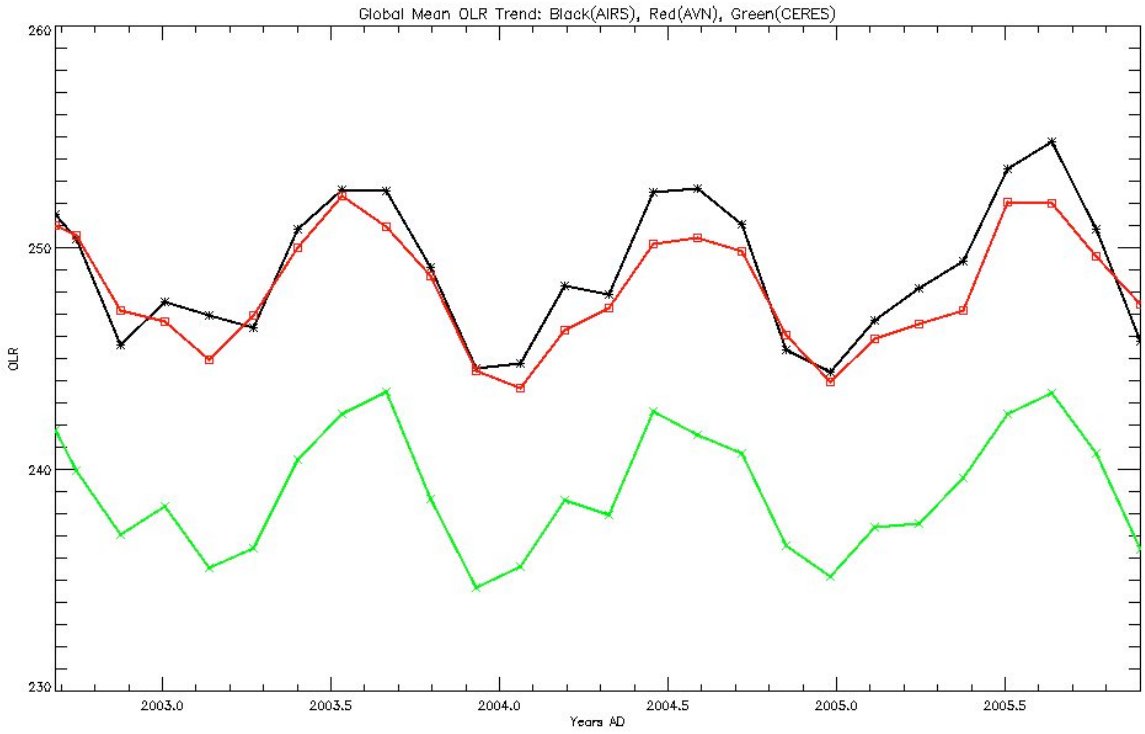

Figure 34 OLR Trend

4.8. Trace Gases

4.8.1. Methane

Parameter List

- **Qual_CH4**
- **CH4_total_column**
- **CH4_VMR_eff**
- **CH4_VMR_eff_err**
- **CH4_eff_press**
- **CH4_verticality**
- **CH4_dof**

Methane is a new research product for AIRS and has not yet been validated. It is a well-mixed tropospheric source gas with a long lifetime (~9 years[1]), and mixing ratio variation is small. However, total column-average mixing ratios as seen by SCIAMACHY indicate that the northern hemispheric column average has roughly 50 – 150 ppbv more CH_4 than the southern hemisphere[2] (out of a global average of about 1700 ppbv), albeit with variations in season and source regions.

AIRS sensitivity to CH_4 is mostly in the mid-to-upper troposphere. Sensitivity is best in the tropics and subtropics, with degrees of freedom of signal (DOFS) varying from 0.9 to 1.3, with DOFS lowered as observations move towards the poles (where DOFS are about 0.6). Preliminary validation indicates an accuracy of 1.2-1.5% depending on altitude[3]. However, as CH_4 is a new research product, more validation will be needed, and there is no previous AIRS benchmark for which to compare the current retrieval. Testing is therefore limited to spot checking to see if the retrieval output is roughly as expected. We compare V5.0.11 results (with included a new transmittance tuning for CH_4) with current results (V5.0.14), the latter including modifications to correct a scan angle dependence on the retrieval.

---

[1] Dentener, F. et al. (2003), Interannual variability and trend of CH_4 lifetime as a measure for OH changes in the 1979-1993 time period, *J. Geophys. Res., 108*(D15), 4442, doi:10.1029/2002JD002916.

[2] Frankenberg, C. J. et al. (2006), Satellite chartography of atmospheric methane from SCIAMACHY on board ENVISAT: Analysis of the years 2003 and 2004, *J. Geophys. Res., 111*, D07303, doi:10.1029/2005JD006235.

[3] C. Barnet presentation at AIRS Science Team Meeting, Pasadena, CA, March 23, 2007.

Figure 35 is an V5.0.11 eight-day map of AIRS-retrieved methane at ~500 mb from August 2005. The mixing ratios and inter-hemispheric differences are roughly about what are expected. The elevated methane over the Atlantic off the west coast of equatorial Africa, and over the Pacific off the west coast of equatorial South America, may not be unreasonable as SCIAMACHY results suggest nearby continental source regions (see Figure 9 of Reference #2). The patchy regions of elevated CH_4 off the coast of Antarctica may not be realistic, and warrant further scrutiny.

Figure 36 compares the September 6, 2002 daytime retrievals at 500 mb using V5.0.11 and V5.0.14. For methane, the largest retrieval algorithm difference is a correction for a scan angle dependence in V5.0.14, the effect of which is most clearly seen in the relative difference in the tropics. While the equatorial mixing ratios are mostly lower by 0 to 2% in V5.0.14, mid-to-high latitudes are higher by about 1 to 4%. Overall, the V5.0.14 retrieval is higher than V5.0.11 by (0.71 ± 0.92)% (1σ std. dev.).

Figure 37 compares V5.0.11 and V5.0.14 retrievals at nighttime from September 6, 2002. The overall morphology of the differences between the versions is similar to Figure 36, however, a comparison of daytime and nighttime retrievals over the Sahara and Sinai deserts indicate that daytime mixing ratios are higher by roughly 100 ppb, for both versions.

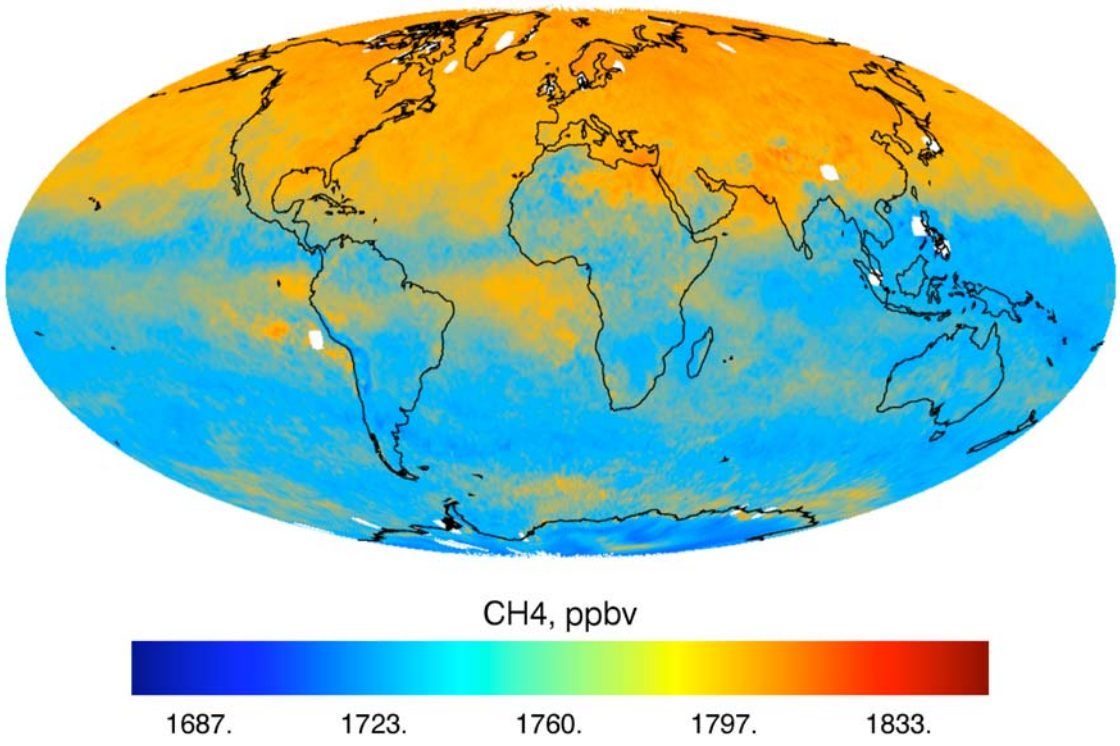

Figure 35 Eight day map of CH_4 at ~500 mb from August, 2005 using AIRS V5.0.11.

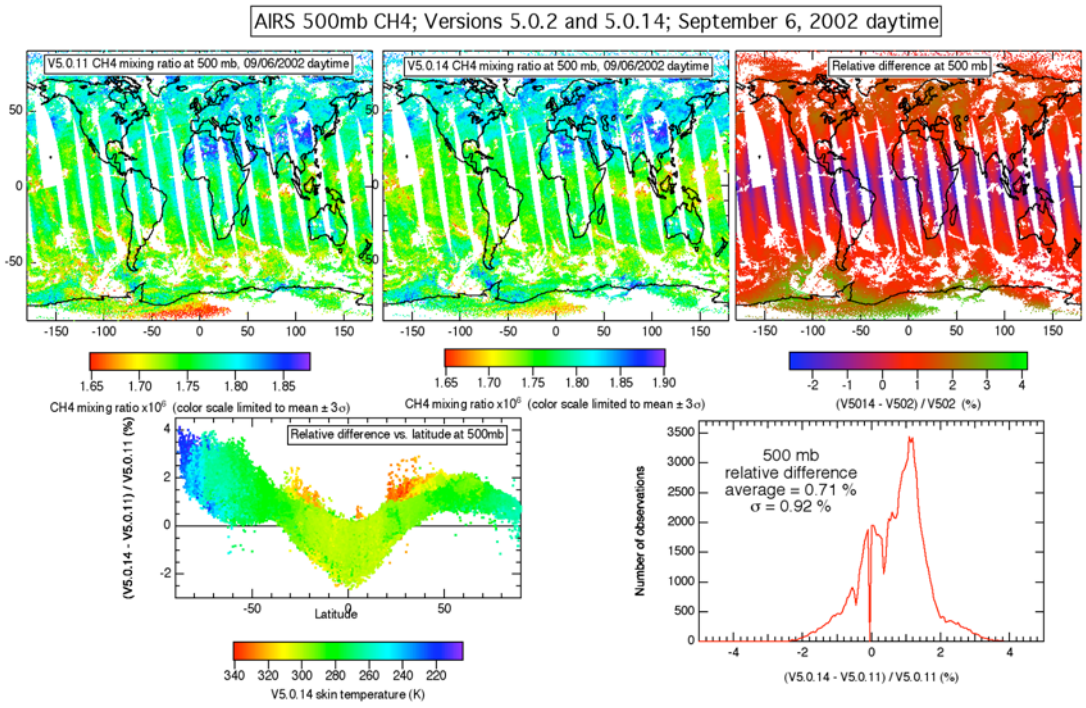

Figure 36 Comparison of daytime AIRS V5.0.1 and V5.0.14 methane at 500mb for September 6, 2002 daytime. The upper left panel shows the retrieved mixing ratio using V5.0.11, the upper middle panel using V5.0.14, and the upper right panel shows the relative difference. The lower left panel shows the relative difference as a function of latitude colored by V5.0.14-retrieved skin temperature. The lower right panel is a histogram of the relative difference.

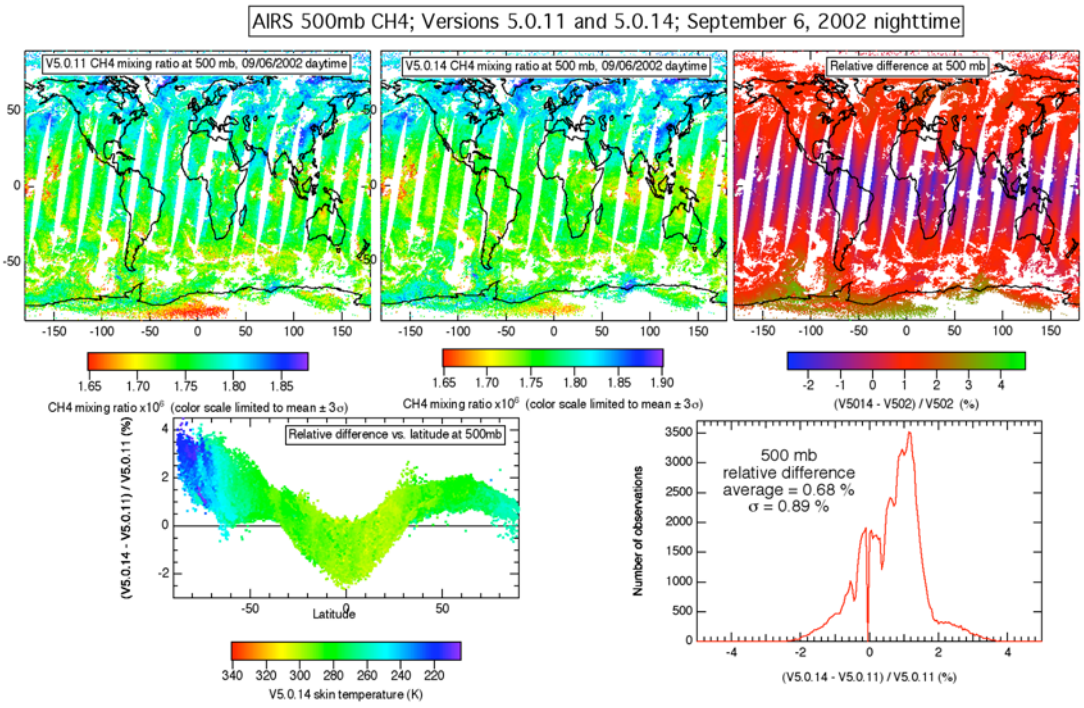

Figure 37 Comparison of nighttime AIRS V5.0.1 and V5.0.14 methane at 500mb for September 6, 2002 daytime. The upper left panel shows the retrieved mixing ratio using V5.0.11, the upper middle panel using V5.0.14, and the upper right panel shows the relative difference. The lower left panel shows the relative difference as a function of latitude colored by V5.0.14-retrieved skin temperature. The lower right panel is a histogram of the relative difference.

4.8.2. Carbon Monoxide

Parameter List

- **Qual_CO**
- **CO_total_column**
- **CO_VMR_eff**
- **CO_VMR_eff_err**
- **CO_eff_press**
- **CO_verticality**
- **CO_dof**
- **CO_trapezoid_layers**

Significant changes have been made for carbon monoxide retrievals between V4 and V5. Retrieval grid trapezoids have been increased from 4 to 9, and the first guess has been changed from the AFGL model to the same first guess used by MOPPITT. Some tuning has been applied to the CO channels, more of which are used in V5. Channels for the CO retrieval are listed below with units of cm^{-1}. Channels newly added for V5 are given in **bold**:

2181.49, 2182.40, 2183.31, 2184.21, 2185.12,
2186.03, 2186.94, 2187.85, **2188.76**, 2189.67,
2190.58, 2191.50, 2192.41, 2193.33, 2194.24,
2195.16, 2196.07, 2196.99, 2197.91, 2198.83,
2199.75, **2200.67**, **2201.59**, 2202.51, 2203.44,
2204.36, **2205.29**, 2206.21, 2207.14, **2208.99**,
2212.71, 2213.64, 2214.57, 2215.50, 2216.44,
2221.12

Preliminary analysis of V5.0.2 results indicated improved sensitivity and accuracy over V4.0.9 [*McMillan presentation at AIRS Science Team Meeting, March 28, 2007*], however, some tuning was made on the AIRS CO channels beginning with V5.0.5. We therefore compare V5.0.14 against V5.0.2 to test changes introduced by the CO channel tuning and possible residual effects due to changes in other parts of the AIRS retrieval since V5.0.2.

Figure 38 compares the CO total column retrievals for V5.0.2 and V5.0.14 for February 28, 2006, daytime. The CO column in V5.0.14 is, by Gaussian fit, higher by 1.6 ± 1.7%, with more variability in this difference around the

mid-latitudes and at temperature extremes. Figure 39 illustrates the mixing ratio difference at the 500 mb level, with the difference being 2.5 ± 2.5 %. Nighttime differences are similar, and a check with another focus day in October, 2005, showed similar differences, as well. Biases for land-only retrievals are slightly lower than ocean-only retrievals. With an estimated 10-15% combined random and systematic error for the CO column [*Wallace McMillan, personal communication*], the bias introduced by the channel tuning is minor.

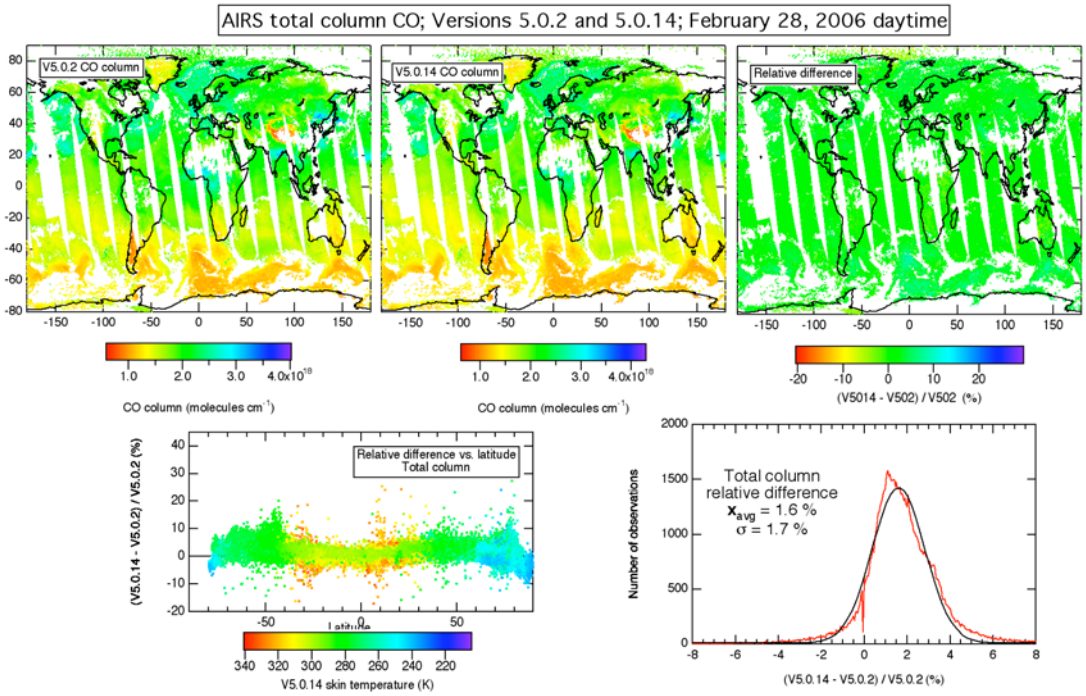

Figure 38 Comparison of AIRS V5.0.2 and V5.0.14 total column carbon monoxide for February 28, 2006 daytime. The upper left panel shows the retrieved column using V5.0.2 the upper middle panel using V5.0.14, and the upper right panel shows the relative difference. The lower left panel shows the relative difference as a function of latitude colored by V5.0.14-retrieved skin temperature. The lower right panel is a histogram of the relative difference (in red) overlaid by a Gaussian fit (in black).

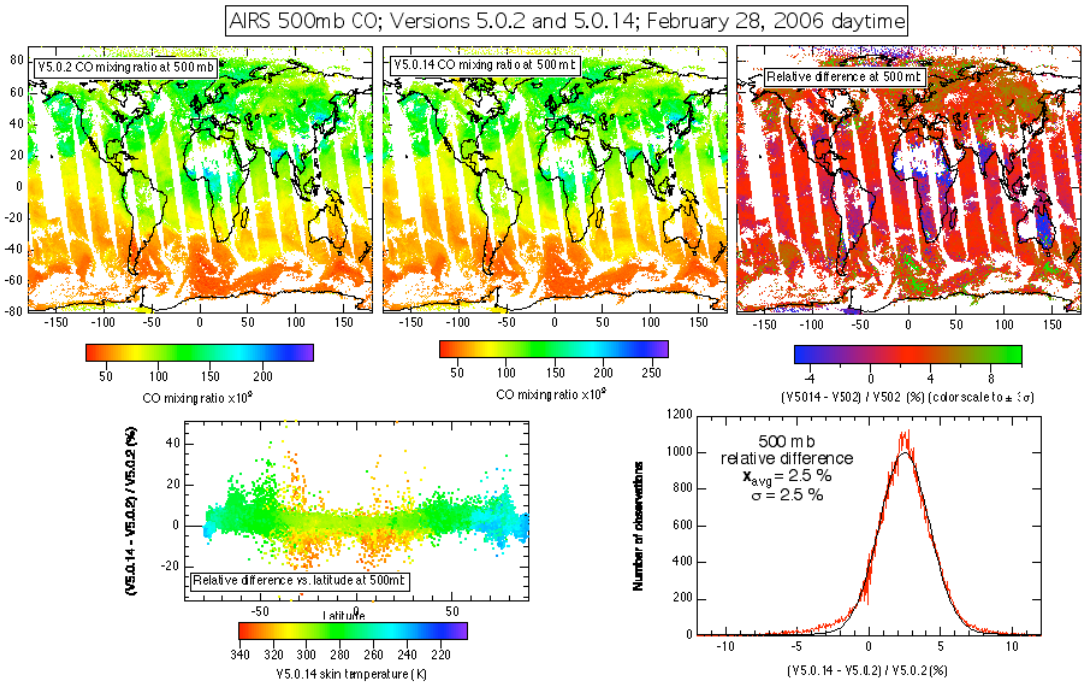

Figure 39 Comparison of AIRS V5.0.2 and V5.0.14 carbon monoxide at 500mb for February 28, 2006 daytime. The upper left panel shows the retrieved column using V5.0.2 the upper middle panel using V5.0.14, and the upper right panel shows the relative difference. The lower left panel shows the relative difference as a function of latitude colored by V5.0.14-retrieved skin temperature. The lower right panel is a histogram of the relative difference (in red) overlaid by a Gaussian fit (in black).

4.8.3. Ozone

Parameter List

- **Qual_O3**
- **totO3Std**
- **totO3StdErr**
- **O3VMRStd**
- **O3VMRStdErr**
- **O3_verticality**
- **O3_dof**

Ozone retrieval methodology for V5 has significantly changed from V4 and as described in *Susskind et al.* [2003], namely in the derivation of the first-guess, channel selection, and "noise propagation threshold" used in the physical (final) retrieval.

In previous versions, the first guess for the ozone profile was regression-based using European Centre for Medium-Range Weather Forecast (ECMWF) ozone profiles as the training set (see *Goldberg et al.* [2003]). Instead of using regression, AIRS V5 uses an observationally-based climatology developed for Version 8 TOMS and SBUV total ozone column retrievals [*McPeters et al.*, 2003]. The climatology is month-by-month on 10° latitude bins. To create a look-up table suited to the AIRS retrieval software, ozone mixing ratios were interpolated by latitude and altitude and converted to slab columns on the AIRS 100-level support grid using the "Partial Column" approximation formula in *Ziemke et al.* [2001]. Where climatological data did not extend the highest or lowest pressure levels of the AIRS support grid, mixing ratio "endpoints" were assumed to extend to such regions.

The channel selection has been extended from 26 (in V4) to 41, and includes the peak of the P-branch in the ozone 10 μm band. The channels are listed below in units of cm^{-1}.

997.11	998.39	999.67	1001.38	1003.53
1005.26	1006.56	1008.30	1010.48	1011.79
1013.11	1014.87	1016.64	1018.41	1020.63
1021.97	1023.31	1024.21	1024.66	1025.11
1025.56	1026.46	1026.91	1027.36	1027.81
1028.26	1028.71	1029.62	1030.07	1030.53
1030.98	1031.44	1061.33	1061.81	1062.29
1063.26	1064.22	1064.70	1065.19	1068.58
1069.07				

Finally, the "noise propagation threshold," ΔB_{max}, discussed in *Susskind et al.* [2003], has been effectively doubled, resulting in less damping of the final profile. Testing for this new scheme involved comparing column retrievals against results from the Ozone Monitoring Instrument (OMI) and profile retrievals against coincident ozonesondes.

Figure 40 shows V4.0.9.0 and V5.0.14.0 results for total ozone on October 7, 2005 daytime[4]. V5 ozone is slightly less in the tropical western Pacific and northern Indian Ocean region, slightly higher in the region south of New Zealand and significantly lower on the Antarctic continent. Figure 41 shows global comparisons to column results from the Ozone Monitoring Instrument (OMI) on the EOS Aura platform [*Levelt et al.*, 2006]. Compared to V4, AIRS-OMI relative biases appear to be slightly worse for V5 in the tropical western Pacific region, and along the coast of Antarctica in regions below Australia and the central Pacific. However the biases in the interior of Antarctica below South America and Africa have been significantly reduced, which is illustrated in the zonal averages for this day shown in Figure 42.

Figure 43 shows the relative difference between co-located daytime total column ozone retrieved by AIRS and OMI binned by latitude during AIRS "Focus Days." AIRS retrievals tend to be within a few percent of OMI, except possibly during the austral spring at high southern latitudes as seen also in Figure 42. Figure 44 shows the average relative bias between AIRS and OMI over the Focus Days shown in Figure 34. AIRS is biased slightly low in the tropics, about the same in

[4] This Focus Day was selected for illustration as it is close to the vernal equinox, facilitating pole-to-pole comparison with the Ozone Monitoring Instrument (OMI) which requires sunlight for its retrievals.

the subtropics, and slightly high in the mid-latitudes and polar regions.

The effect on retrievals of switching from a regression-based to a climatological first guess at several sites is illustrated in Figure 45, where the left panels show AIRS-sonde biases using regression as a first guess, while the right panels show such biases using climatology as a first guess. (AIRS V5.0.7 was used for this illustration, however spot testing indicates negligible column differences between V5.0.7 and V5.0.14.) Using a climatology reduces biases in the mid to lower troposphere where AIRS has little sensitivity. Stratospheric biases are about the same. Upper tropospheric biases have been reduced in tropical regions, but somewhat worse in mid-latitude regions. Validation of V5 AIRS ozone against coincident ozonesondes, aircraft measurements and ground-based Dobson spectrometers is ongoing.

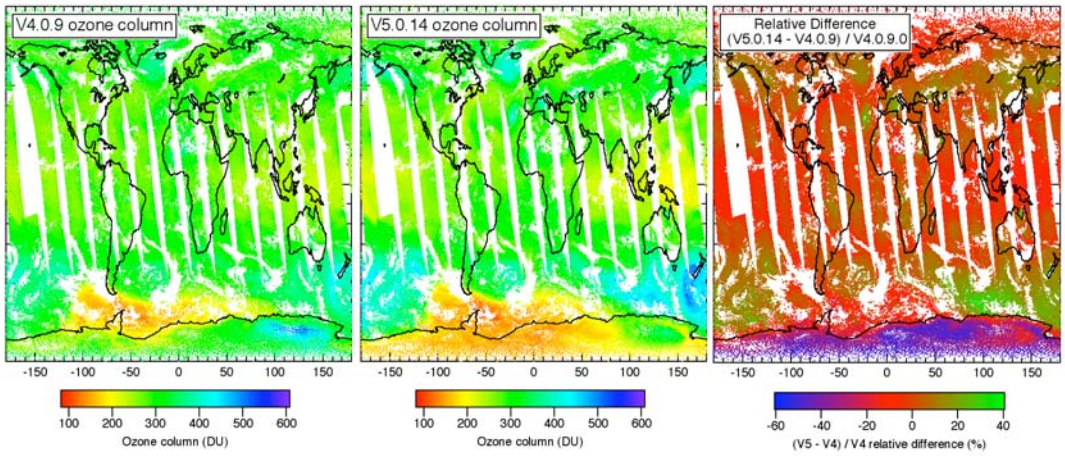

Figure 40 Total column ozone retrievals from V4.0.9.0 (left panel) and V5.0.14.0 (middle panel), and relative difference (right panel) for October 7, 2005 daytime. Only retrievals with a Qual_O3=0 flag are shown.

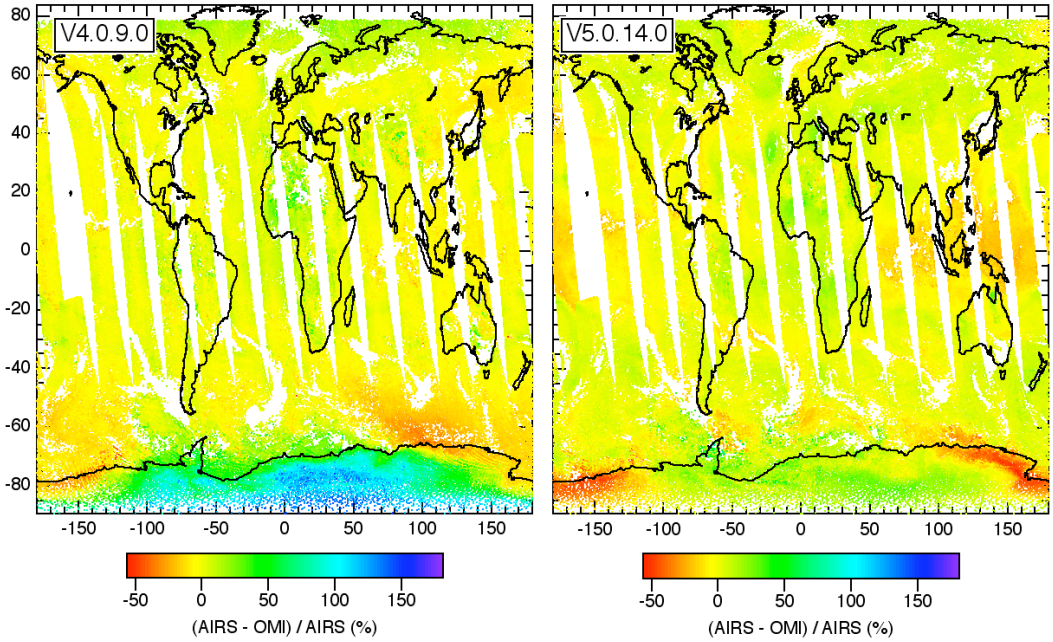

Figure 41 Relative difference between AIRS and OMI on October 7, 2005 using V4.0.9.0 (left panel) and V5.0.14.0 (right panel). AIRS was compared to OMI where the geographical center of an AIRS retrieval was within a 0.25°x0.25° gridbox of the OMI high-resolution L3 product.

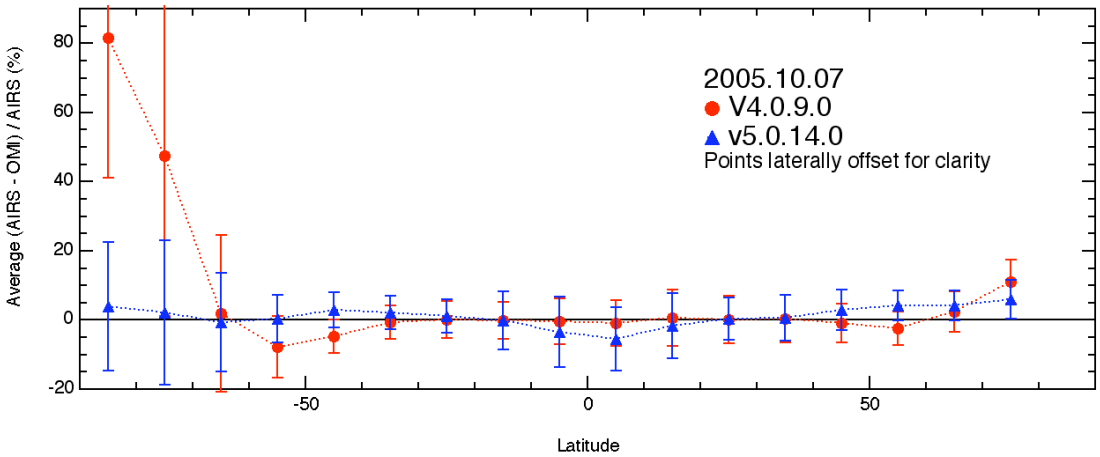

Figure 42 Average relative difference between AIRS and OMI for daytime October 7, 2005, binned by latitude. Error bars are 1σ standard deviation.

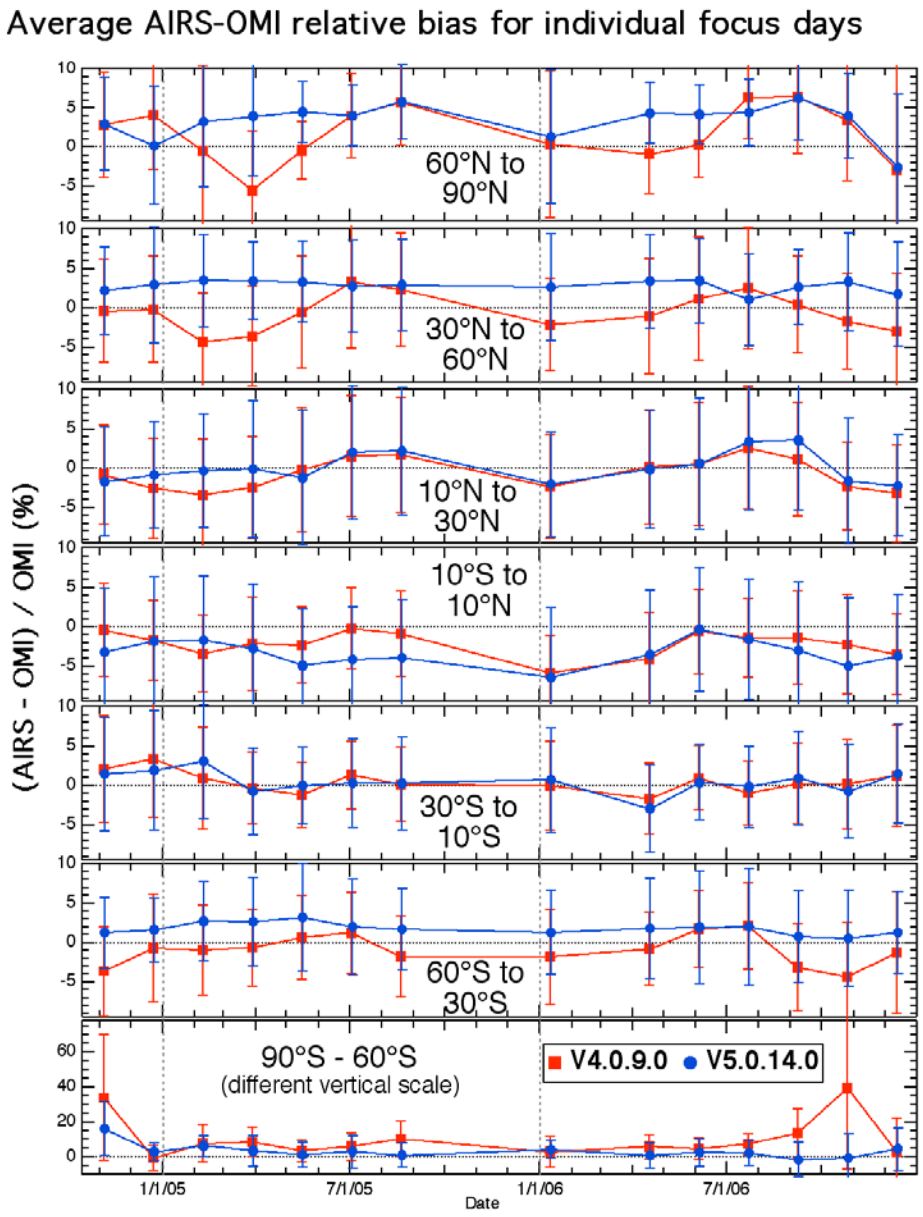

Figure 43 AIRS-OMI relative difference vs. date, binned by latitude. Column retrievals were compared where the center of an AIRS retrieval fell within a grid box of the high-resolution (0.25° x 0.25°) OMI Level 3 product. Error bars are 1σ standard deviation. Note that as OMI relies on backscattered UV for its measurement, comparisons do not extend into the polar night.

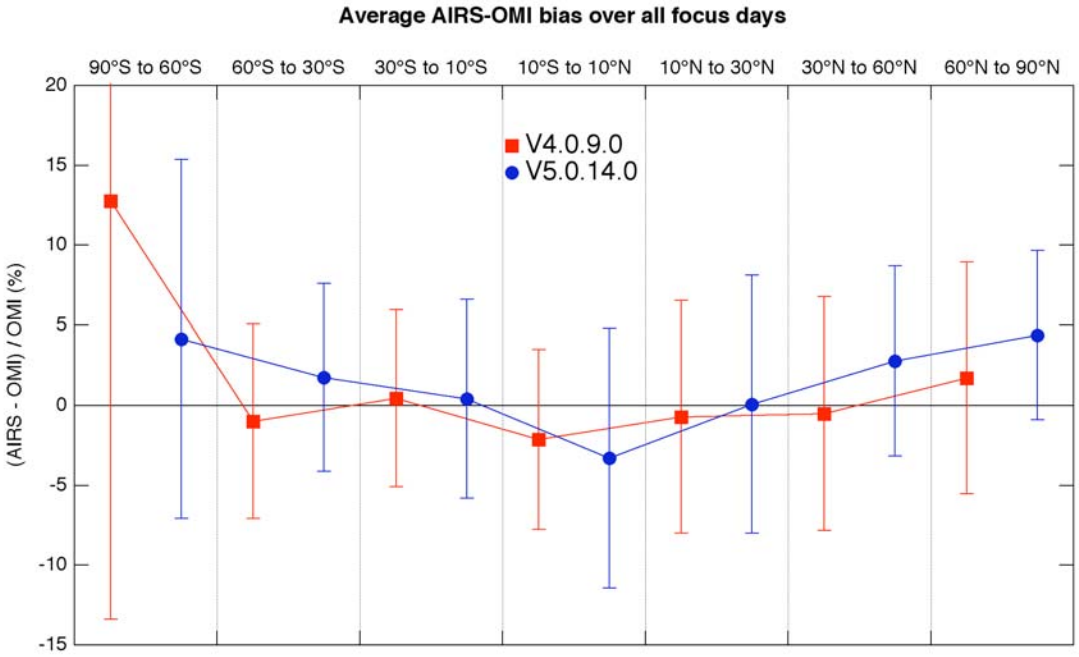

Figure 44 Average relative difference between AIRS and OMI total column ozone over the Focus Days shown in Figure 43. Error bars are 1σ standard deviation.

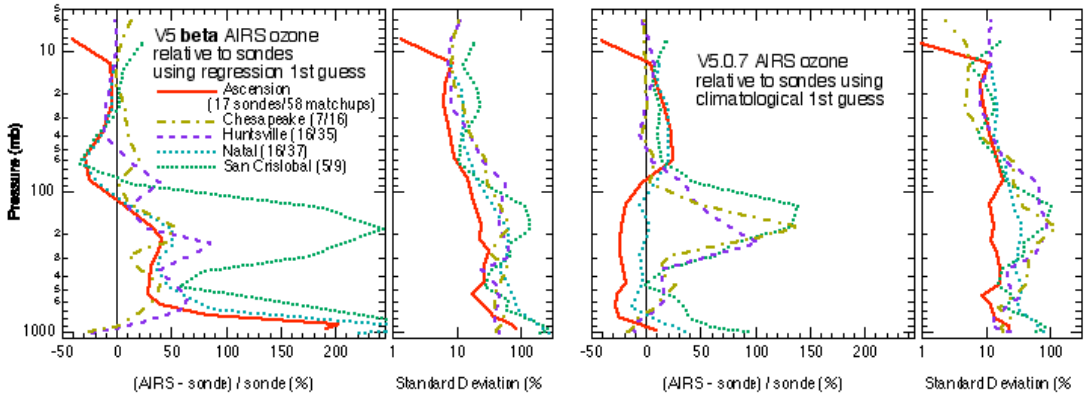

Figure 45 Average relative differences and standard deviations of AIRS retrievals to ozonesondes with regression first guess (left panels) and climatological first guess (right panels). AIRS-sonde matchups occurred with AIRS observation within 3 hours and 50 km of sonde launch. Sonde profiles were smoothed to AIRS standard vertical layer depths prior to comparison.

4.9. Emissivity Parameters

Parameter List

- **Qual_Surf**
- **numHingeSurf**
- **freqEmis**
- **emisIRStd**
- **emisIRStdErr**

The parameter **Qual_Surf** indicates the quality of the surface emissivity and is indicates increasing reliability with decreasing value; 0 is highest, 1 is acceptable and 2 is considered unreliable. The fractional number of retrieved states with **Qual_Surf** unacceptable should decrease in V5. Figure 46 shows histograms of **Qual_Surf** probability conditional on surface type (ocean and land) for a day of V4 and V5. Over land, more states have higher quality in V5, but none of them have the highest quality. Over ocean the number of unacceptable states has decreased, but the number of highest quality states has also decreased. **Qual_Surf** has modestly improved in V5, but it is a mixed result.

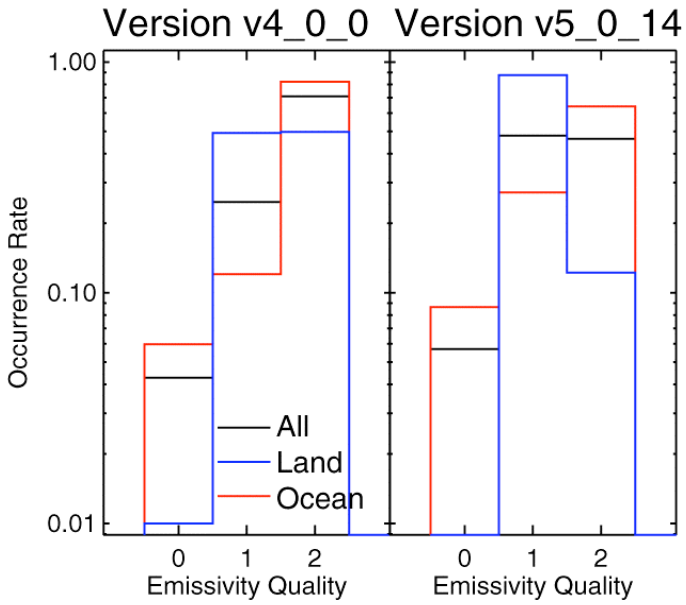

*Figure 46: Fractional Number of states with **Qual_Surf** equal to 0, 1 and 2 for V4 (left) and V5 (right). Ocean scenes are red, land scenes are blue and all scenes are black. Data is all states retrieved on 11 Nov 2006.*

The parameter **numHingeSurf** is the number of emissivity hinge points. **FreqEmis** is the frequency of the hinge points and both vary between states. of the V5 goals was to regularize the set of reported hinge points. V4 and 5 products contain 89 and 86 unique hinge points, but each stated contains at most 49 or 48 hinge points.

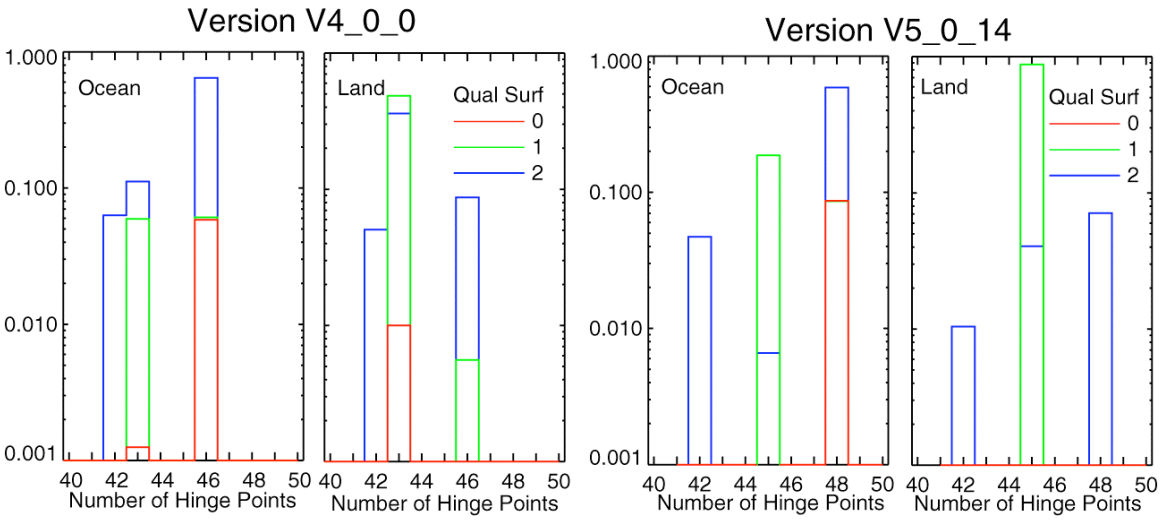

Figure 47: Histograms for the number of reported hinge points reported from V4 (left) and V5 (right) products for: all states (red), acceptable (green) and highest quality (dotted black). Data is from 11 Nov 2006.

Therefore there is considerable variability in the location of the hinge points. Figure 47 shows occurrence histograms of **numHingeSurf** for V4 and 5 products binned by **Qual_Surf** and separated by surface type. Figure 48 shows occurrence histograms of actual hinge points The highest quality states in the V4 products have either 43 or 46 hinge points whereas the corresponding states in the V5 products have a fixed set of 48 hinge points. States with acceptable quality in the V5 products still has two distinct sets of hinge points. Finally four different sets of hinge points are occur for states with unacceptable quality. Although modest improvements have been realized in reducing hinge point variability, the current hinge point products remain unnecessarily complicated.

The parameter **emisIRStd** is retrieved over ice and land, but derived from an emissivity model over ocean with a four-hinge-point correction for model error. The ocean emissivity model has not change from V4 and testing is restricted to land states. Five test sites listed in Table 2 cover a wide range of surface conditions of soil and vegetation cover. For each site, the Johns Hopkins Emissivity Database was queried for sets of samples classified as: “grass”, “carbonate”, “basalt”, “quartz” or “deciduous.” Mean and standard deviation spectra were derived from each set and these were used to establish test metrics

at each site. There were two test requirements; first that the V5 emissivities should agree more with the database emissivity spectra than the V4 products, and secondly that 75% of the derived spectra should lie with 1-standard deviation of the mean spectra.

Table 2 Emissivity Test Sites

Index	Site	Latitude	Longitude	Soil Composition Vegetation Cover	# DB Samples
1	ARM-CART SGP	36.62º	-97.5º	Grass / felsic soils	2
2	HaNegev, Israel	30.94º	34.97º	Carbonate soils	48
3	HaGolan, Israel	33.03º	36.04º	Grass / basaltic soil	34
4	Eygpt-One	27.12º	26.10º	Quartz sand	29
5	Salonga NP, Congo	-2.5º	21.2º	Rainforest	1

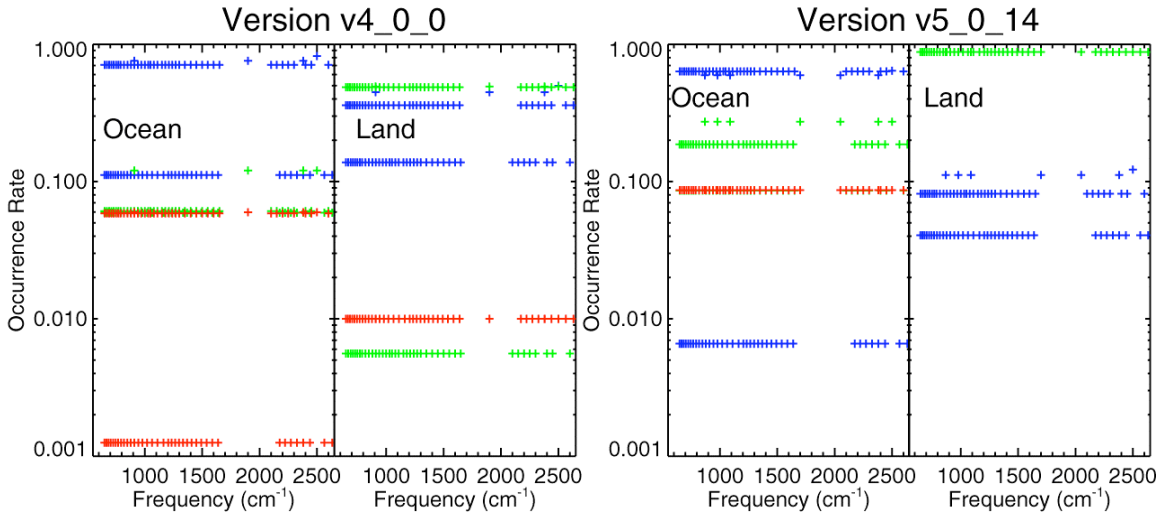

Figure 48 Histogram of emissivity by frequency. Red, green and blue are ***Qual_Surf*** *equal to 0, 1 and 2 respectively.*

Figures 49 compares the retrieved emissivity spectra with the database spectra at each of test sites. The database mean emissivity spectra with 1σ error bars

are shown by the magneta curves (the set of grass spectra from the database contains only one sample and does not have error bars.) The comparisons are from data collected on 16 Nov 2002, each row is for each of the sites (1-5) with the closes observation on the day and night passes shown on the left and right sides respectively. Comparisons are shown for 4 sets of data. The black lines are the V4 products (labeled V4_0_0x). V5 can be run in three different modes: a nominal AMSU/AIRS mode shown by the green curves (V5_0_7x), an AIRS/AMSU/HSB mode shown by the blue curves (V5_0_7h) and a AIRS only mode shown by the cyan curves (V5_0_7s) curves. The processing software was modified after the test data was created and the nominal mode was rerun and is shown by the red curve (V5_0_14x). The differences between the V5_0_14x and V5_0_7x in all comparisons are small compared either to the variability within the database subset or the differences between V5 and V4 products.

The upper row shows comparisons at the ARM-CAR SGP site. We expect day and night emission spectra to be the same and they are except for a long wave and short wave offsets. Because the spectral structure is added via a regression, we see that the regression introduces consistent emissivity spectra for the day and night overpasses. The final retrieval corrects the regression-derived emissivity spectra with long and short wave offsets, and these are not consistent on the two overpasses. Also we note that the short wave emissivity spectra is significantly lower in the V5 products and is more consistent with database emissivity spectra

The second row shows emissivity spectra at the Egypt-one test site. This site is covered by dunes of quartz sand with virtually know vegetation. Quartz emission spectra contain a doublet at 9 μm, which the V4 product either failed to identify or placed at the wrong frequency. The V5 emissivity spectra show the doublet or a reasonable approximation to it, but do not show a Christensen frequency near 1350 cm^{-1} (high emissivity). This error in the emissivity occurs in the water vapor sounding band and could be correlated with errors in retrieved total precipitable or near-surface water vapor. Lastly, we note differences in the emissivity spectra between day and night overpasses --the 11 μm doublet is less accurately resolved in the day overpass and the short-wave emissivity is 10% higher in the night overpass. However these differences are small compared to the improvement from V4 to V5.e third row shows emissivity spectra over the uplifted flood basalts of HaGolan – a region of weathered flood basalts covered by grasses and shrub vegetation. The day and night emissivity spectra differ by around 10% at shorter wavelength and like the Egypt one site, the nighttime overpass has higher emissivity than the daytime overpass. The retrieved emissivity falls outside the expected range at the 1350 cm^{-1} Christensen

frequency. However 75% of the spectra lies within database variability and therefore the retrieved emissivity spectra meets requirements.

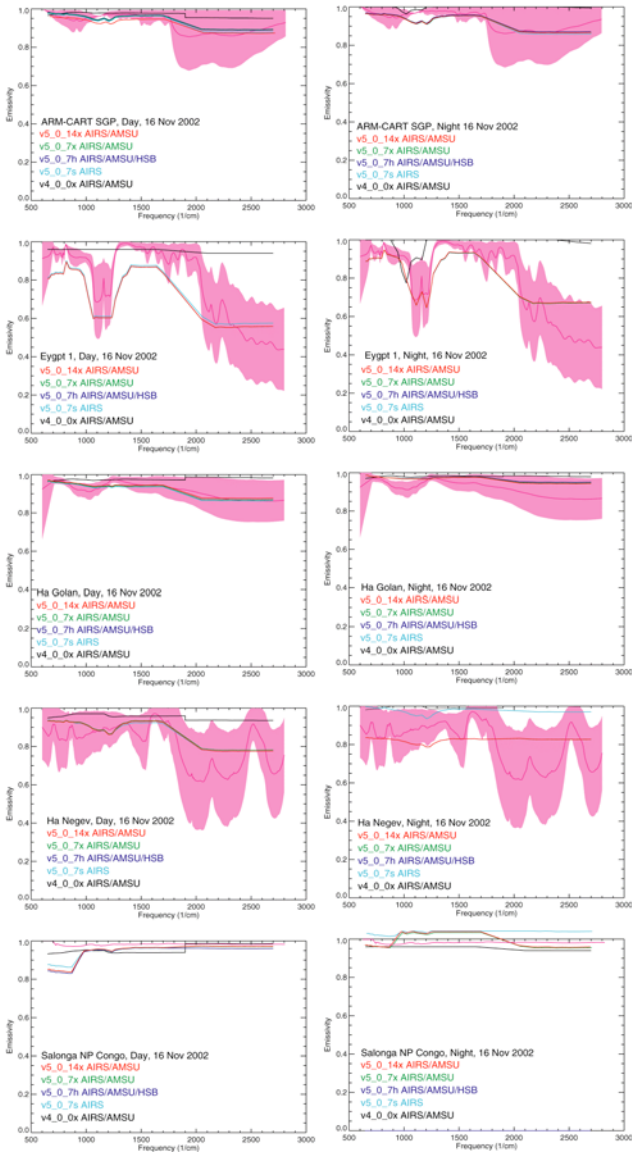

Figure 49: Retrieved emissivity spectra at ARM-CART, SGP; Egypt-1, HaGolan, HaNegev and Salonga emissivity test sites versus laboratory/field measured emissivity for grassland, quartz sand, basalt, carbonate and deciduous forest. (See test for further detail).

The fourth row compares retrieved emissivity spectra at Ha Negev. The region is desert with soils weathered primarily from carbonates. The nighttime case is the

single case where the V5_0_14x product is significantly different from the V5_0_7 products. The V5_0_14x shows less day – night differences at short wave, but about 10% lower long wave emissivity during the night. The emissivity database has large variability in emissivity between samples and the retrieved emissivities lie with the loose requirements even though the characteristic carbonate emission features long wave of 900 cm^{-1} are not present.

The bottom row shows the comparisons at the Salonga National Park in the Congo. This is a region of dense pristine tropical rainforest. The observed radiances were highly cloud contaminated and the retrieved emission spectra varied by about 5% depending on product type. Day and night spectra were similar although the nighttime AIRS-only product (V5_0_7s) and V4 products were larger than one (capped by a 1.05 maximum limit constraint). Generally the V5 products are worse than the V4 product and are greater than 1 as often (or more in the AIRS only product). The limit test could not be implied to this case because the database does not provide an variability estimate, but the uncertainly in the emissivity of highly vegetated tropical scenes is sufficiently well know to conclude that the V5 products do not pass at this site.

Figure 50 shows histograms of the reported emissivity errors for the V5 product – the V4 product did not report any errors. From the standpoint of testing, the presence of errors in the V5 product pass the testing requirement, but there are several reasons to believe these reported errors do not characterize the actual errors in the products. First, most histograms have a spike at 0.01 which is several order of magnitude larger than the surrounding distribution (the vertical scale is logarithmic) and indicates that emissivity error is dominated by only the lowest values regardless of whether the reported emissivity quality is 0, 1 or 2. Secondly the histograms are not significantly different in regions of the spectra where the atmosphere is opaque (e.g. 650, 1550 and 2000 cm^{-1}) versus regions where it is transparent (e.g. 870 cm^{-1}.). Lastly over land, the errors for the highest quality products (quality 1) are not significantly different from those over water, even though the ocean first guess is much better constrained by the initial first guess. So in summary, the reported emissivity errors although improved, appear unreasonable.

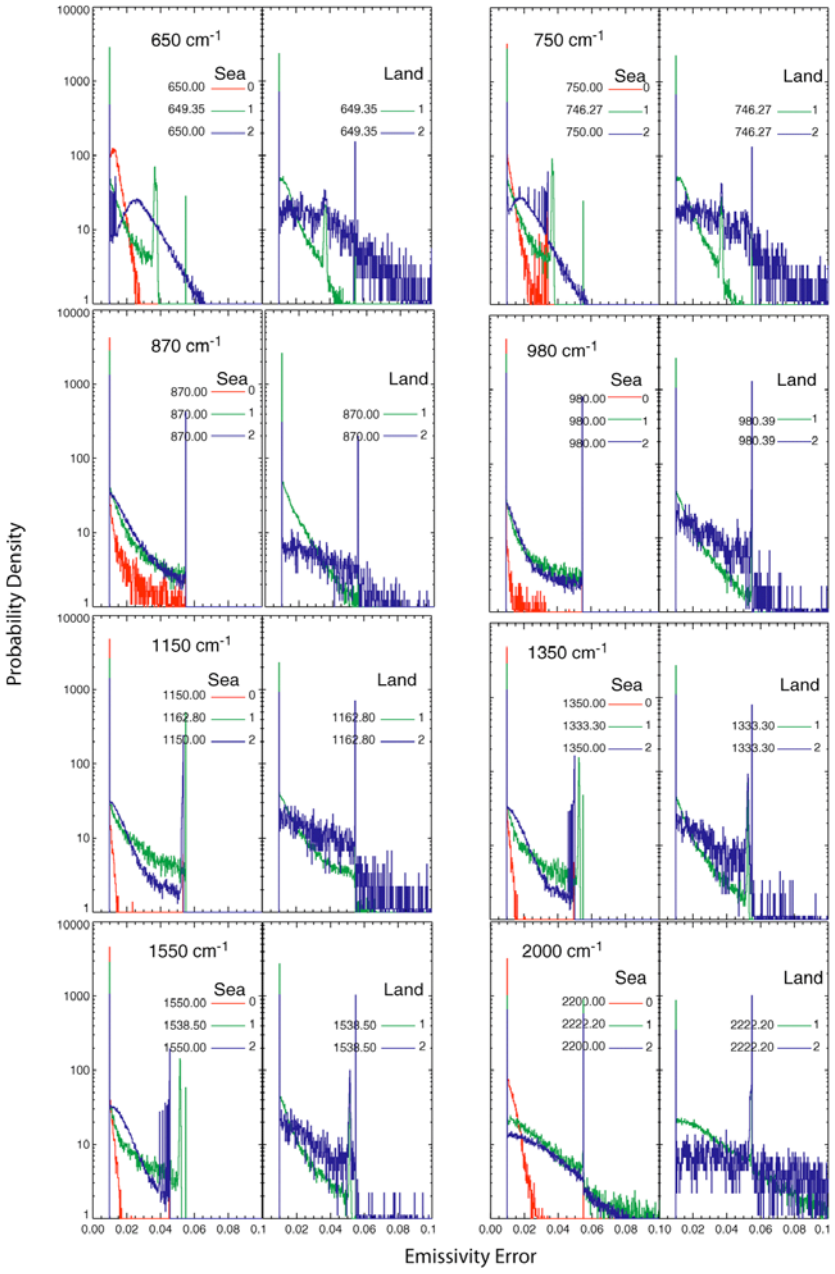

Figure 50: Histograms of reported emissivity error density near 8 frequencies (left to right, top to bottom, 650, 750, 870, 980, 1150, 1350, 1550 and 2000 cm^{-1}. Left and right side of each panel are data over sea and ocean separated by emissivity quality, 0 (red), 1(green) and 2 (blue). Near each frequency, the histogram is constructed from the hinge point containing the largest number samples within the five closest and the frequency of that hinge point is shown in the legend.

4.10. Tropopause Pressure, Temperature, Height

Figure 51 shows a pole-to-pole, northward-moving (daytime) slice of tropopause properties during the half-orbit centered on 20:47 on September 6, 2002.

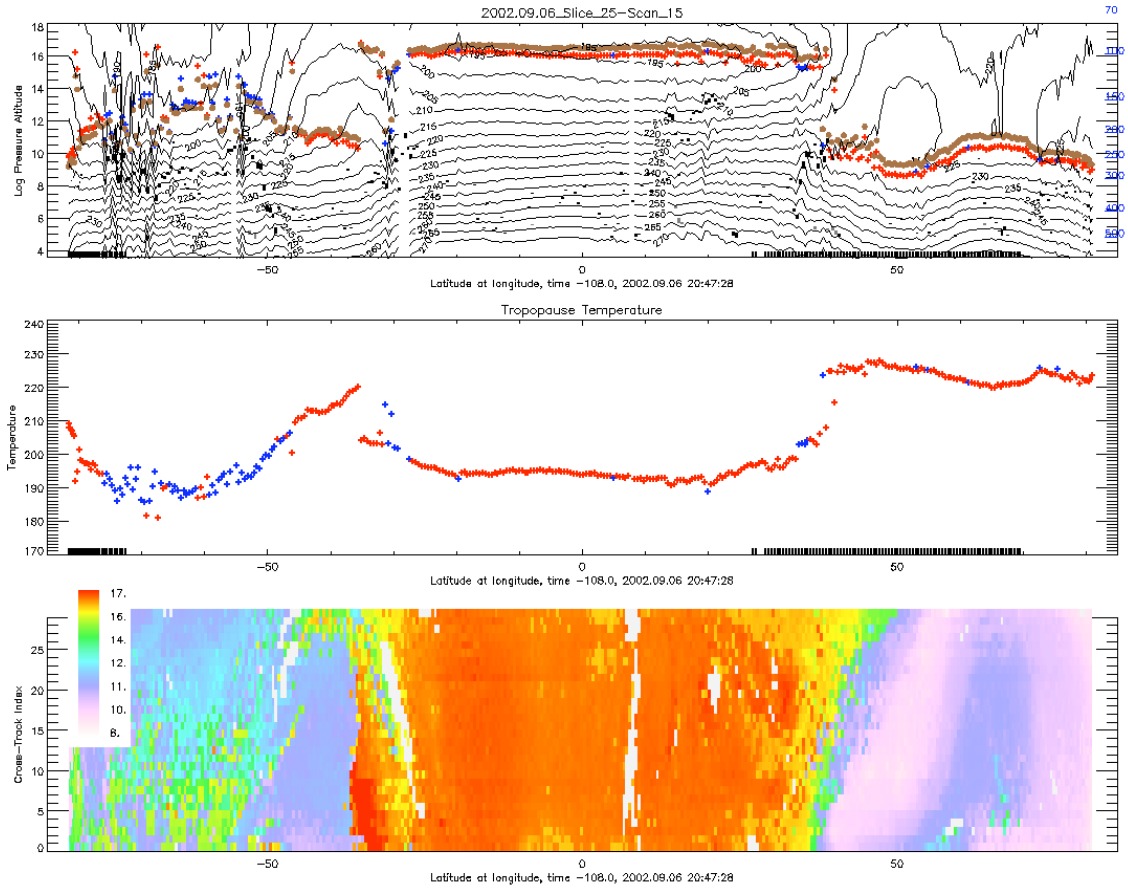

Figure 51 Half-orbit centered on 2002.09.06 20:47 at scan 15 of 30. Top panel: Contour plot of retrieved temperatures, with tropopause log-pressure altitudes as plus signs (blue is best retrieval value down to tropopause, PBest > Ptropopause*; red is poorer retrievals) and tropopause geopotential height as filled dots. Black rectangles represent clouds, with their vertical extent proportional to cloud fraction. The small vertical lines at bottom represent land. Middle panel: Retrieved tropopause temperature versus displacement along the orbit, same color scheme as top panel. Bottom panel: Retrieved geopotential heights for the same orbit, plotted versus along-track and cross-track indices. Missing data are poorer quality retrievals (*PBest > Ptropopause*).*

An additional 31 half-orbits were examined on this day. Figure 51 gives a typical example of the AIRS retrieved tropopause structure. All the quantities examined

appear reasonable, and exhibit properties consistent with known behavior of the tropopause. The exception is a scan angle-dependent striping in tropopause quantities. This striping does not appear to dominate over geophysical variability, though exceptions are seen. This is discussed below.

The top panel of Figure 51 shows a cross section of temperature versus log-pressure altitude (scale height = 7 km) for scan 15 of 30 in the orbital swath. Also show in the top panel are the log-pressure altitude of the tropopause as plus signs (calculated from *PTropopause*), and the tropopause geopotential altitude as brownish dots (from *GP_Tropopause*). The slight discrepancy between them comes from the use of a fixed scale height in calculating log pressure altitudes. Both quantities show the sudden drop in tropopause height when moving to higher latitudes. This represents the location of the jet stream; the increased cloudiness there is consistent with midlatitude storminess. The constant tropical tropopause altitude of ~16 km is typical of most AIRS cross sections on this day, as is the higher variability at higher latitudes. As seen in Figure 51, the tropopause generally lies above region of lapse rates of ~10 K/km, consistent with standard definitions of the troposphere. The largest local variability in tropopause heights is seen at highest latitudes, apparently due to increased retrieval noise there, though some high latitude variability is likely associated with geophysical variability. This will be addressed in validation comparisons in the future. Note that the regions of higher variability from noisier retrievals are associated with regions of greater cloud cover, as represented by the black rectangles in the figure.

The middle panel of Figure 51 shows tropopause temperature (*T_Tropopause*) versus latitude for scan 15 of 30. These temperatures correspond to the altitudes and heights in the upper panel. The lowest tropopause temperatures are found in the tropics, with a warmer (and lower) tropopause at higher latitudes. The transitions between tropical and mid-latitude tropopause are obvious.

The bottom panel of Figure 51 shows the horizontal variability of tropopause geopotential height (*GP_Tropopause*). This horizontal variability is broadly consistent with the properties shown in the upper panel of Figure 51, including a higher tropical tropopause rapidly transitioning to a lower polar tropopause. Missing data represent retrievals where the lowest good retrieval is above the tropopause (*PBest* < *PTropopause*); these are retrievals shown in red crosses in the upper panel. These footprints are excluded to give an idea of how frequently tropopause retrievals are of lower quality.

Figure 52 shows tropopause geopotential heights for a single orbit centered at 12:35 UTC on September 6, 2002, but with the color bar range limited between 13 and 17 km. The two panels represent best retrievals (top) and all retrievals

(bottom). Along-track striping is apparent in Figure 52, especially around 25 S and 25 N. This structure is roughly symmetric around the centerline of the swath. Figure 52 shows a worst-case example for September 6, 2002, and most orbits are less severely affected than this. Further analyses will be needed to determine if this striping affects the climatology of tropopause properties derived from AIRS.

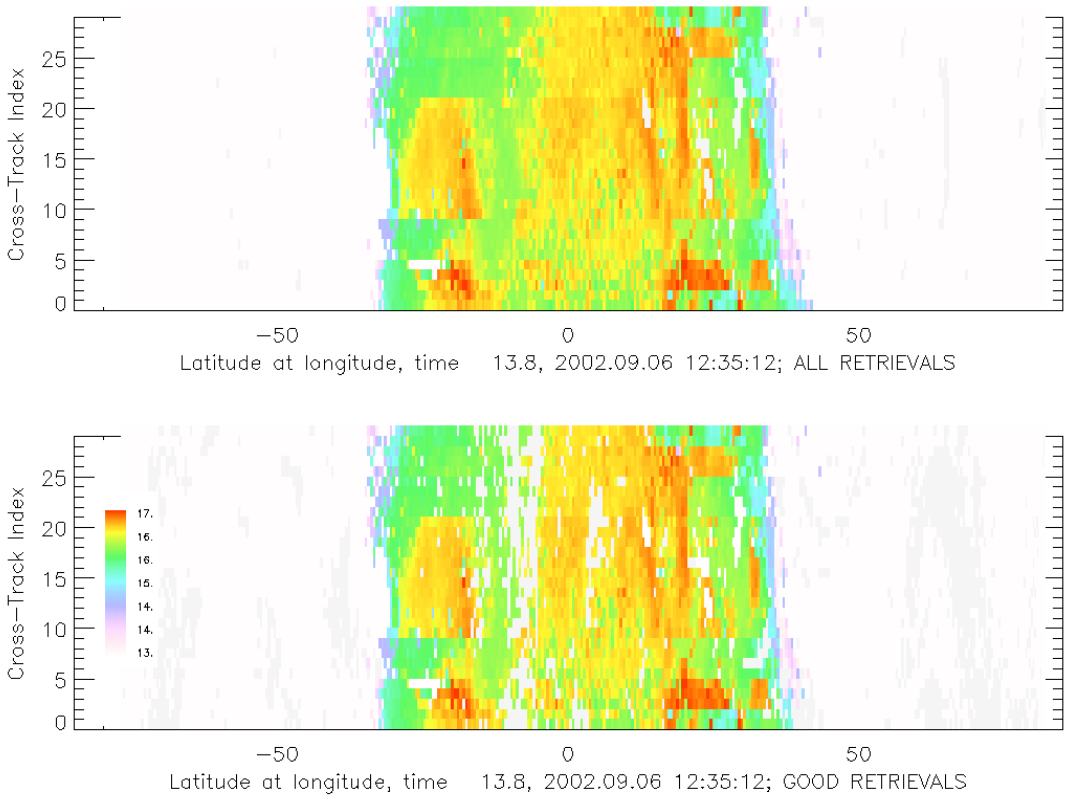

*Figure 52 Retrieved geopotential heights for the orbit centered on 12:35 on September 6, 2002, versus along-track and cross-track indices. The color bar is limited to 13 to 17 km to highlight the along-track striping. Top panel: All retrievals. Bottom panel: Good retrievals only (*PBest > Ptropopause*).*

Figure 53 shows all tropopause pressures versus latitude for September 9, 2006. The preference for certain pressures is apparent. A preliminary check revealed no obvious scan-angle dependencies to the banding in Figure ZZ. The pressure banding is reflected in the tropopause geopotential heights, though heights are

not fixed to discrete increments but instead show latitudinal dependence of the bands. This is consistent with how heights are calculated.

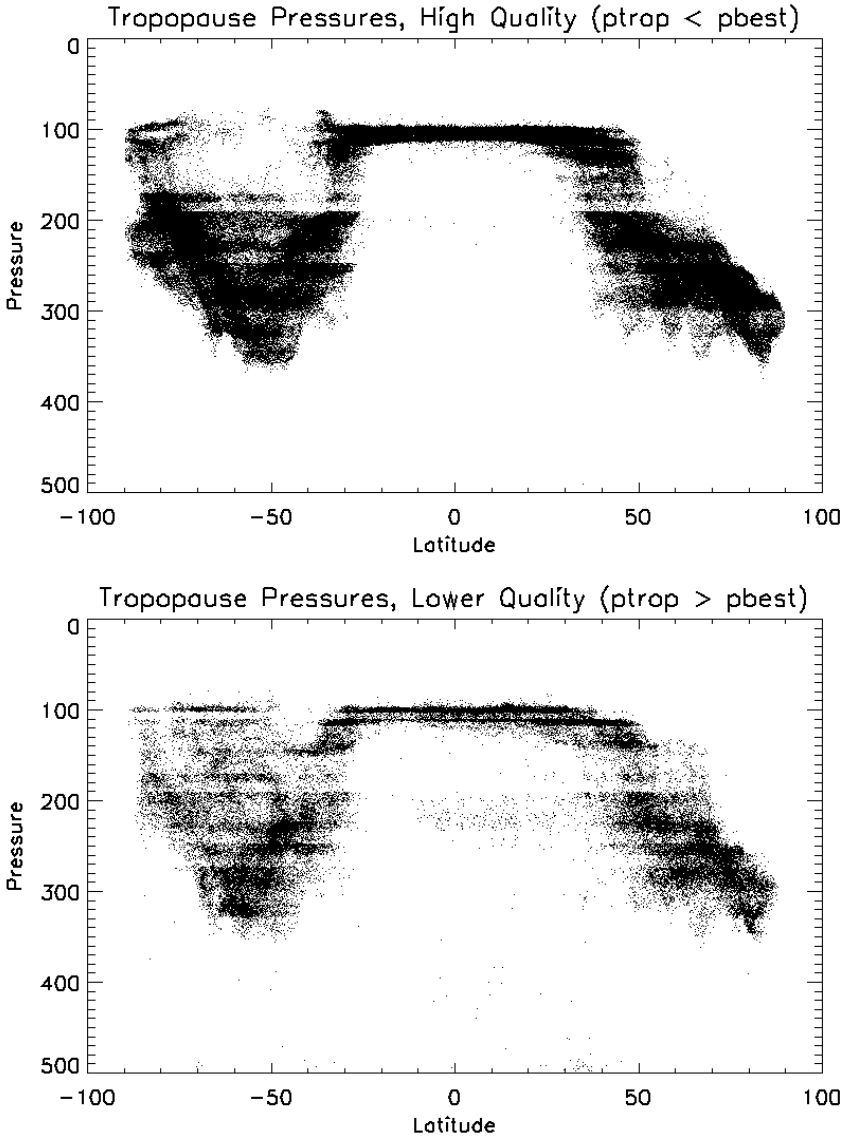

Figure 53 Tropopause pressure versus latitude for highest quality retrievals (upper panel) and lower quality retrievals (lower panel).

4.11. Geopotential Height and Surface

Parameter List

- **GP_Height**
- **GP_Height_MWOnly**
- **GP_Surface**

There is no requirement on the accuracy of these parameters in the ATLP. Nevertheless, the **GP_Tropopause** shown in Figure 52 appears reasonable.

4.12. Additional QA parameters and flags

4.12.1. RetQAFlag

See the limit test document.

4.12.2. initial_CC_Score

The **initial_CC_score** is the principal component score from the initial regression. It is the RMS difference between the cloud cleared radiances and the reconstructed radiances in units of NeN. Since the regression is trained on high quality cloud cleared radiances, the CC score is a measure of quality of initial cloud clearing. The results of initial regression is rejected if CC score is larger than 10.

The initial effective noise amplification factor is another measure of quality of initial cloud clearing, which is used by the initial regression. However, the correlation coefficients of these two parameters indicate pretty low correlation, 0.30, on September 6, 2002 data.

There is some indication of striping in the image of the initial cc score. The Figure 54 is the map of the CC score for granule 43 of October 2, 2006. The scale was chosen to magnify the striping. The figure clearly shows striping. It could be the result of using channels that have deteriorated progressively since the regression channels were selected at the beginning of the mission. Channel selection for initial regression and local angle correction should be revisited for V6.

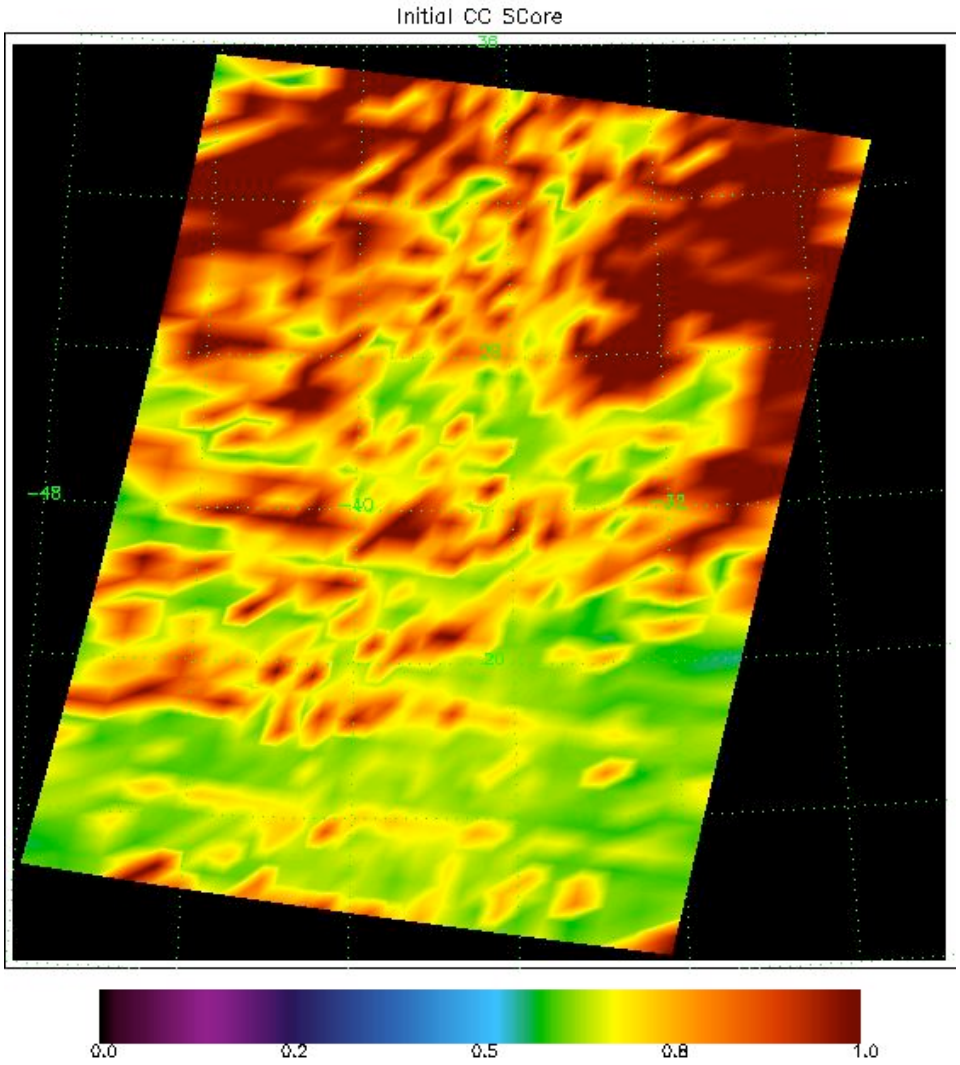

Figure 54 Map of Initial CC Score for granule 43 of Oct 2, 2006.

4.12.3. retrieval_type and IR_Startup

The retrieval type indicates the program flow. The retrievals of type 100 mean that MW only algorithm failed to retrieve high quality profiles, usually over coastline or over precipitating cloud. Since the MW rapid forward algorithm used by the AIRS level 2 PGE cannot account for ice cloud or precipitating cloud in the profile, the MW only algorithm rejects these profiles. This parameter has been in place since the earliest working version of the level 2 PGE software and does not require extensive testing. However, this parameter is deprecated in favor of quality flags introduced in V4 and modified in V5.

IR_Startup is a parameter that indicates the first guess to the initial cloud clearing. It has the value of 0 for the MW only algorithm (AIRS+AMSU, old baseline), the value of 1 for cloudy regression (AIRS Only), or the value of 2 in the blended system where we run both MW only algorithm and cloudy regression (AIRS+AMSU, new baseline). Sample granules were checked to make sure that the document is correct. In the current implementation of level 2 PGE, the value of this parameter is not footprint dependent. So it could have been implemented as an attribute.

4.12.4. dust_flag and dust_score

The dust score and the dust flag was developed at the University of Maryland at Baltimore County. The deficiency of the original V4 algorithm was discovered by JPL staff members and the modifications were developed together. The parameters were computed in the AIRS level 1 PGE and then copied to level 2 output. The threshold value for the dust flag is set to 381. The map of dust flag shows some false alarms over convective cloud.

It has been tested on Asian dust as well as Saharan dust with great success. Although this is not a validated data product and there is no dust retrieval implemented yet, this dust flag was successfully used to track Asian dust to the west coast of the United States. It may have come out of the east coast of US near Long Island, NY.

Figure 56 is the map of a dust score for a sample granule near Canary Island in the tropical Atlantic Ocean. The dust score does not work well over land or cloud. But the map clearly indicates dust storm in the lower one third of the granule, which is confirmed by the Vis/NIR map in Figure 56.

The series of maps of dusty footprints for the focus days indicates that the dust flag is quite conservative, and that the false alarms are scattered and quite scarce. Figure 57 is a sample dust map for July 14, 2003.

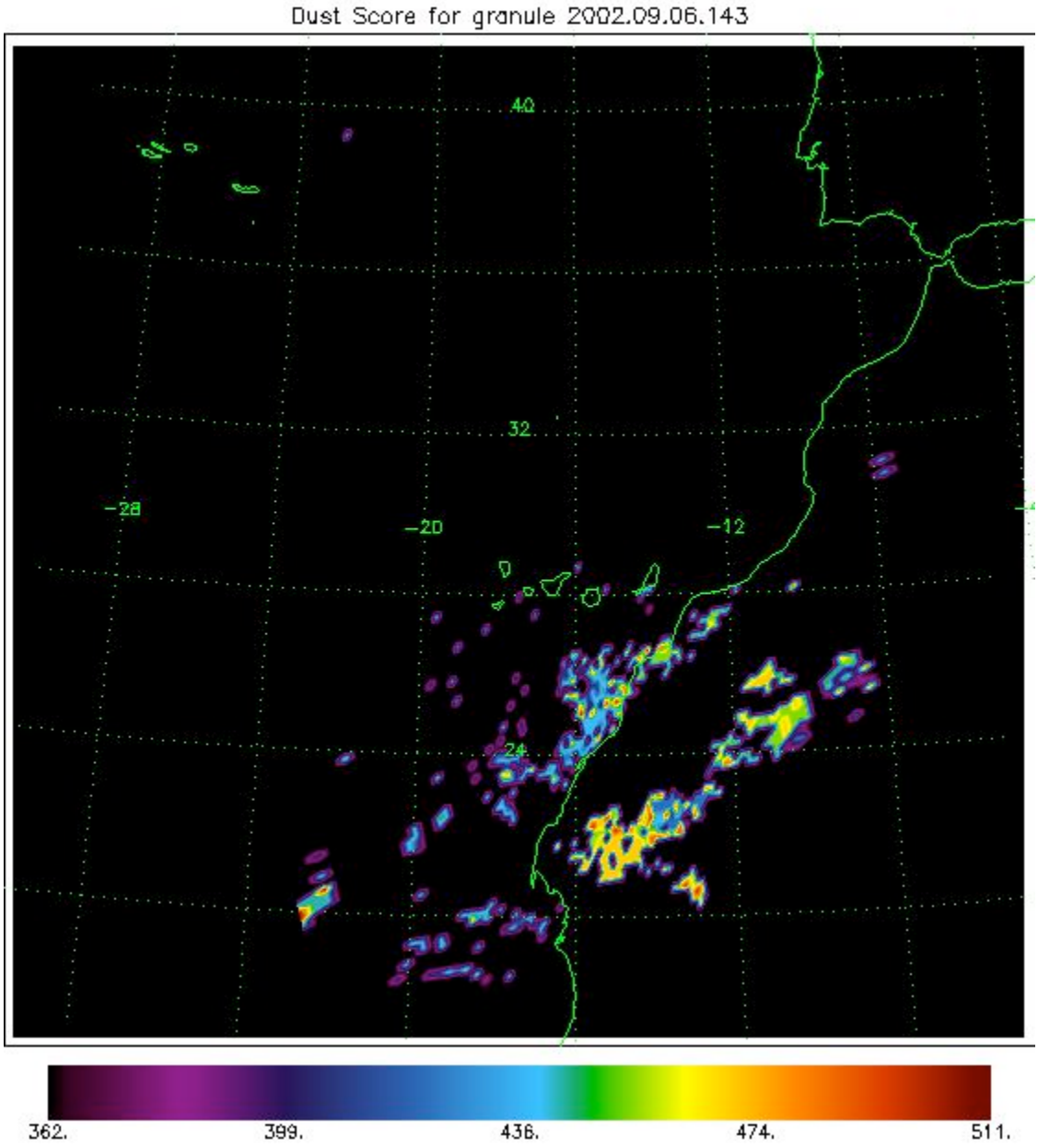

Figure 55 Map of Dust Score for granule 143 of Sept 6, 2002

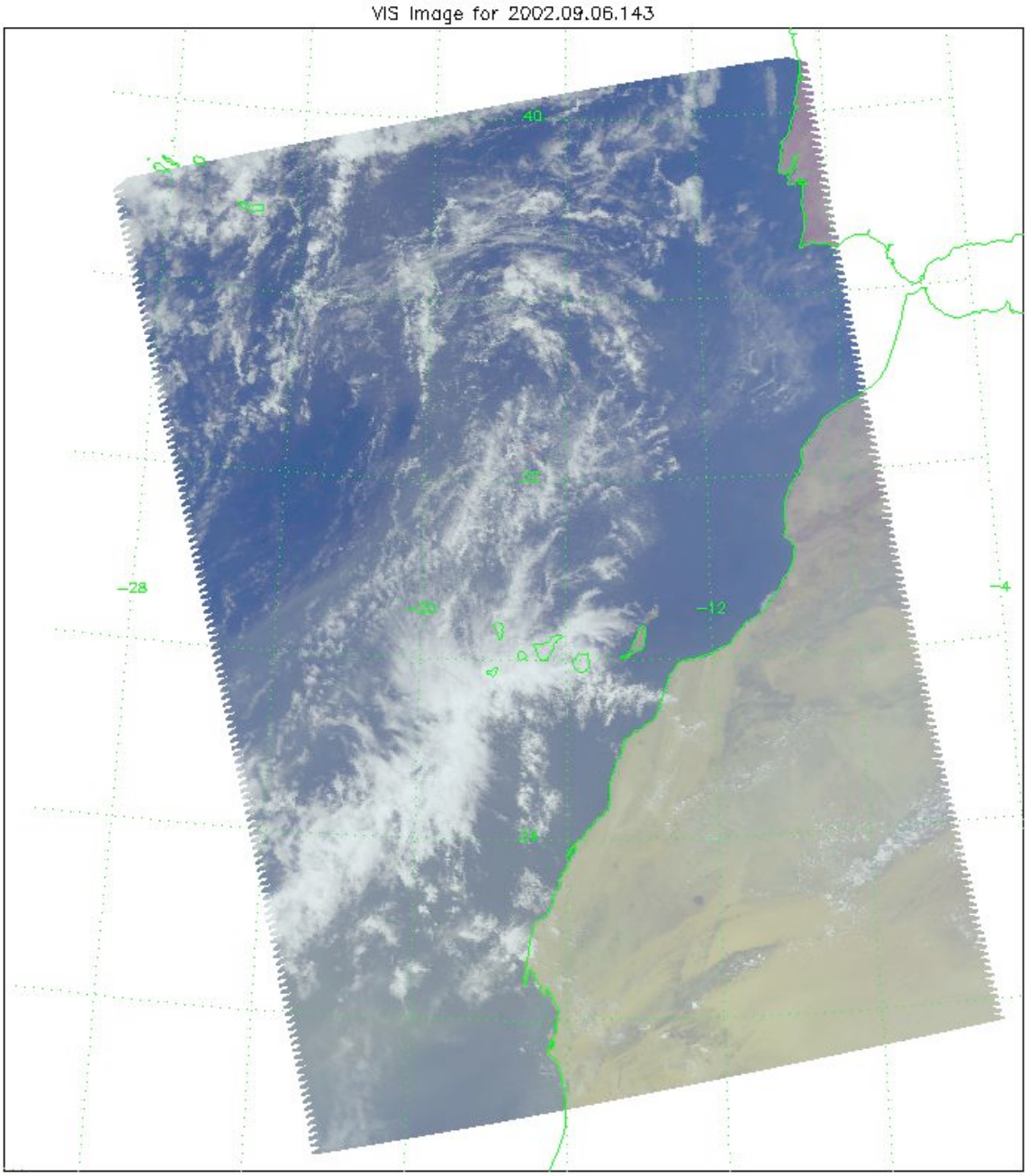

Figure 56 Vis/NIR Image of the same granule.

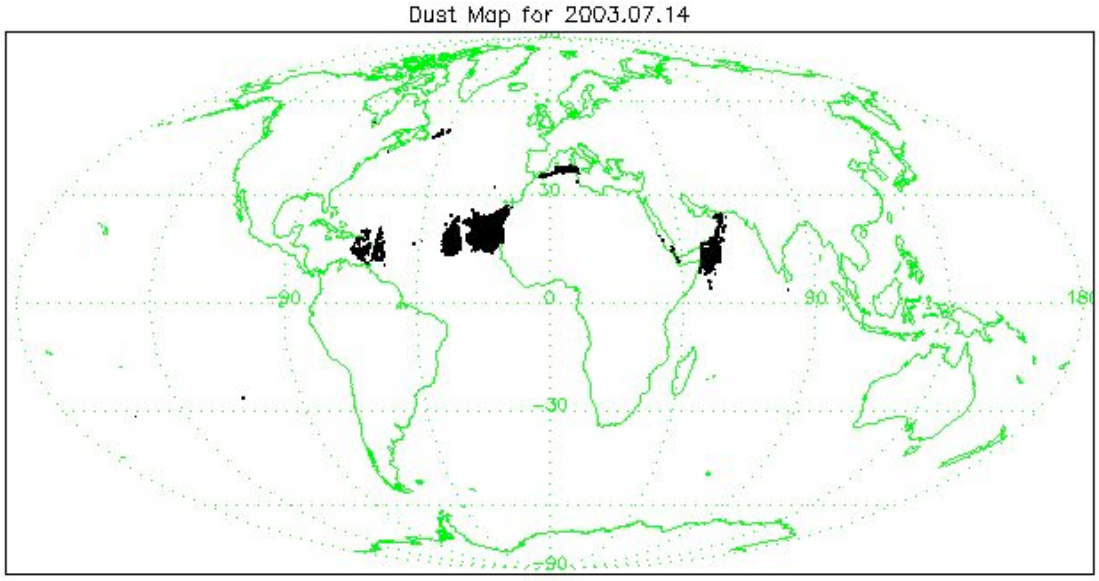

Figure 57 Dust Map for July 14, 2003.

4.12.5. all_spots_avg, CC_noise_eff_amp_factor, and CC1_noise_eff_amp_factor

These parameters are folded into the error estimates.

4.12.6. SO2 Brightness Temperature Difference and SO2 Flag

The brightness temperature difference and the SO_2 flag were introduced in early part of V5 development. Two channels were chosen so that their response to water vapor is quite similar, but one is in the middle of 1350 cm-1 SO_2 band and the other is away from the SO_2 band. The brightness temperature difference has the global mean of about 0.5K and the standard deviation of about 1K. But the brightness temperature difference becomes much smaller over an SO_2 plume, due to the sensitivity of the first channel to the SO_2. The SO_2 flag is turned on when the brightness temperature difference gets smaller than -6. This is a very conservative limit to reduce the number of outliers. Over relatively clear sky and away from polar regions, -3 or -4 could have been suffice to detect SO_2 plumes. Also the sensitivity of the channels to water vapor is strong enough that a low level (in height) SO_2 plume cannot be detected by this simple

algorithm. The number of footprints where the SO_2 flag was turned on in a granule was put into the metadata and hence can easily be searched through tools supplied by ESDIS.

The flag and the brightness temperature difference are computed in the AIRS level 1 PGE. The parameters are then copied to the level 2 support output by the level 2 PGE. The correct transfer of the data from level 1 to level 2 support files was confirmed using a sample granule.

The flag has been operational in the NOAA Near Real Time system since January of 2006 and has been working very well. There were more than half a dozen SO_2 warnings that were confirmed by volcanic ash advisory and other news releases. No false alarm has been issued since the software was modified to ignore footprints immediately before and after spacecraft maneuvers.

Figure 58 shows a map of SO_2 brightness temperature for granule 66 of May 21, 2006. The AIRS observation was within 24 hours from the dome collapse of Soufrier Hills Volcano in Montserrat, which released the SO_2 plume. The comparison of the figure with the map of retrieved total SO_2 column from UMBC shows remarkable resemblance. The color palette was reversed to show the resemblance.

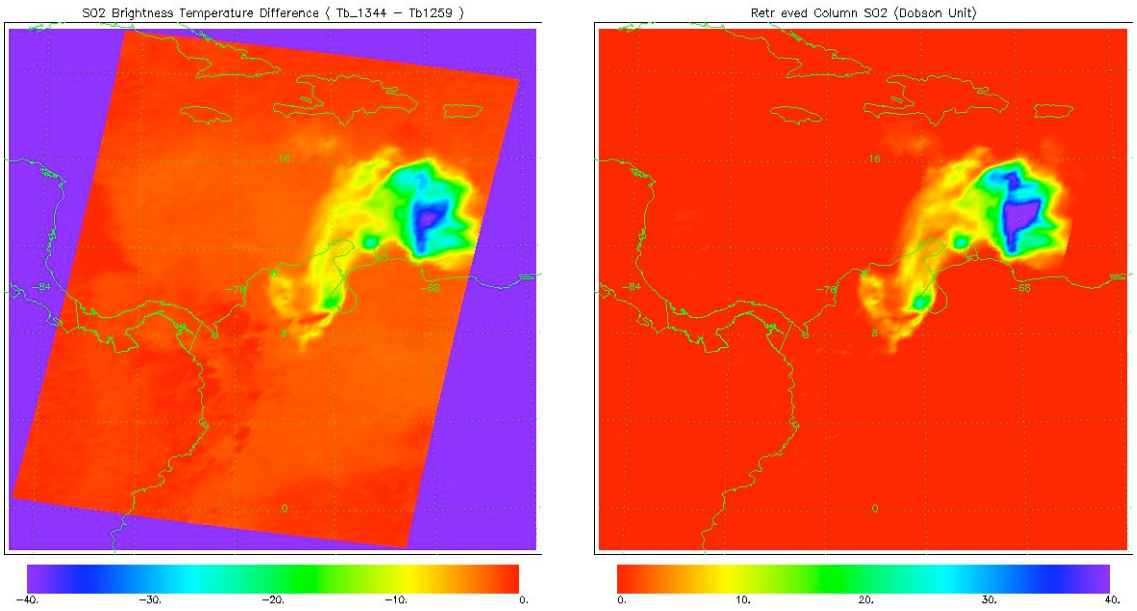

Figure 58 SO2 Brightness Temperature Difference (left) and Retrieved Total SO2 Column (right)

It has been observed that the initial PC score gets very large over a volcanic SO_2 plume and most such profiles are rejected, although the physical retrieval algorithm should not be affected by the existence of SO2 in the atmosphere. This does not happen often enough to modify the level 2 PGE at this point. We estimate that a granule a month will be affected by this symptom.

4.13. Cloud Cleared Radiances

Cloud cleared (CC) radiances are the estimated radiance from the clear region of the scene and are intermediate product in the retrieval process. CC radiances are tested using calculated radiances derived from ECMWF forecasts and the retrieved states. The testing measures are that the CC radiances are 1) unbiased and their errors should be characterized by the reported CC radiance errors. The Student's T-test and Chi-Squared test provide statistical significance of bias and variance.

The following notation is used in this section. A CC radiance is represented as I^{CC}, a calculated radiance is represented by I^{calc} and the CC radiance error is represented by σ; radiances have dimensions of brightness temperatures.. The radiances depend or frequency v, and statistics are calculated over ensembles of footprints or states *N*. The index *i* refers to states (footprints) within the ensemble. The Student's T-test τ is define as

$$\hat{o} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \left(\frac{I_i^{CC} - I_i^{Calc}}{\sigma_i} \right).$$

If I^{CC} and I^{calc} are sampled from the same distribution, that is each samples the same state, then in the limit of large N the T-test is distributed according to a normal distribution with a mean of zero and a standard deviation of one. If τ is less than –2 or greater than +2, the confidence is 95% that the calculated and cloud-cleared radiances are not derived from the distributions. The normed chi-squared is defined as

$$\chi^2 = \frac{1}{N} \sum_{n=1}^{N} \left(\frac{I_i^{CC} - I_i^{Calc}}{\sigma_i} \right)^2,$$

and is distributed along a chi-squared distribution which approaches a normal distribution with mean equal to one and standard deviation equal to $N^{-½}$ *in the limit of large N. The* χ^2 *statistics assesses the normalcy of the residuals* $I^{CC} - I^{calc}$ *and accuracy with which the error* σ_i *characterized the residual*.

Figure 59 shows the t-Test derived from the CC radiances compared to calculated radiances from V4 and V5 solutions and from the ECMWF analyses. Radiances of surface sensing channels are not reliable from the ECMWF calculations over land because the emissivity is assumed equal to one and the reflectivity is zero – similarly daytime short wave radiances are also inaccurate over ocean because the sea surface is assumed Lambertian. Across the entire spectrum, the t-test is significantly reduced in V5 as compared with V4. Much of this improvement in V5 arises because of the larger reported CC radiances errors, but these are not the errors used internally in the minimization. The T-tests from the V5 products in the 15 and 4 μm temperature sounding channels are also much better than those derived from the ECMWF calculated radiances. However, the T-tests in the water vapor sounding channels at 6 μm are not significantly better in the V5 derived t-test. Similarly, the t-tests values in the 10 μm ozone-sounding channels over ocean are also not much different in the V5 and ECMWF derived t-tests even though the ECMWF ozone is biased more than the water vapor product. The fact that the t-test is uncorrelated to the quality of the ECMWF calculated radiances suggests that the observed biases are not related to biases in the CC radiances and that the radiances calculated from the solutions are fitting errors in the CC radiances.

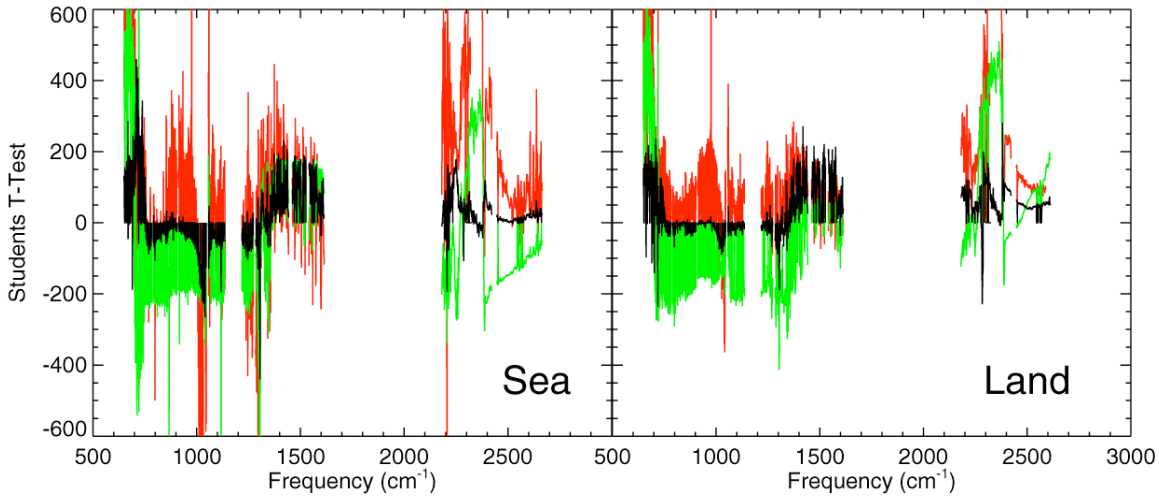

Figure 59: T-test statistics derived for CC radiances from 11 Nov. 2002. The green curve are derived using differences between V5 CC and calculated radiances from ECMWF. The black curve is T-test using V5 CC minus calculated from V5 solutions; the red curve is the same as the black, but for V5 CC and solutions.

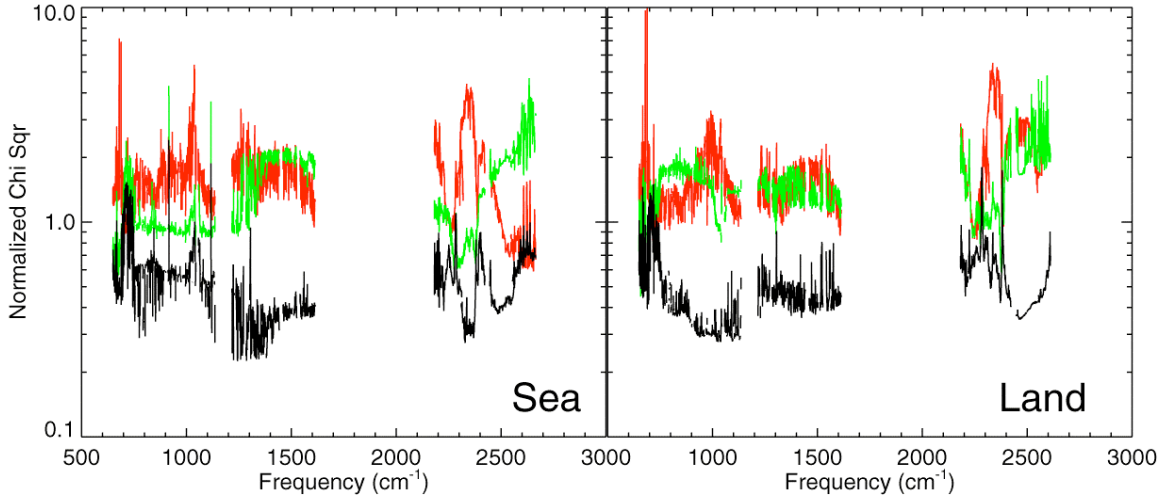

Figure 60: χ^2 statistics derived for CC radiances for 11 Nov 2002 as in Figure 59.

Figure 60 showing the chi-squared statistic characterizes the accuracy of the CC radiance errors. The chi-square shown in Figure 60 is the adjusted chi square, $\chi'^2 = \chi^2 - \tau^2/N$, and removes the bias component of the chi-square that is already characterized by the t-test. The χ'^2 for V4 product is generally less than 2; the χ'^2 for the V5 product is around 0.5 and the χ'^2 calculated from ECMWF calculated radiances is closest to one. The 3-σ confidence interval for these χ'^2 are around 0.1 (thousands of observations) and it is highly likely that the CC radiance errors underestimate the errors in V4 and overestimate the errors in V5. Furthermore, because the χ'^2 derived from ECMWF-derived radiances is close to one, especially in long-wave surface-sounding channels over ocean, the reported V5 CC radiance errors probably include error contributions from the ECMWF radiances used in training the error mode. Although the CC radiance errors are different in V5, they are not better and may actually be worse because they have contributions from the ECMWF analysis. Also the a major component of improvement in the t-tests arise because the CC radiance errors are now larger than the errors in V4. Accordingly the V5 CC radiances and their errors are not improved and fail these tests.

4.14. Temperature and Water Vapor Trends

Divakarla et al. (2006) identified an increasing trend (~0.1 K/year) in the air temperature bias in the lower troposphere with the V4 retrieval algorithm. The V5 algorithm implemented a time varying CO_2 amount to correct this trend but it still exists in the V5 algorithm (Figure 61). So the trend is likely not due to CO_2. The trend is also seen in the Microwave retrieval and the AIRS-Only retrieval. Interestingly, The bias trend over land is in the opposite direction (Figure 62) and the yield for the best quality retrievals based on the air temperature quality flags is decreasing. Figure 63 shows that the surface temperature bias over land also has a decreasing trend of ~1 K/year. Since there may be similar trends in the ECMWF data, validation investigations should be performed to determine whether the apparent trend is due to the AIRS data or ECMWF.

Figure 64 shows that there is also a decreasing trend in the upper tropospheric water vapor mixing ratio with respect to ECMWF over, water, land, and ice. There are also significant seasonal variations in the bias.

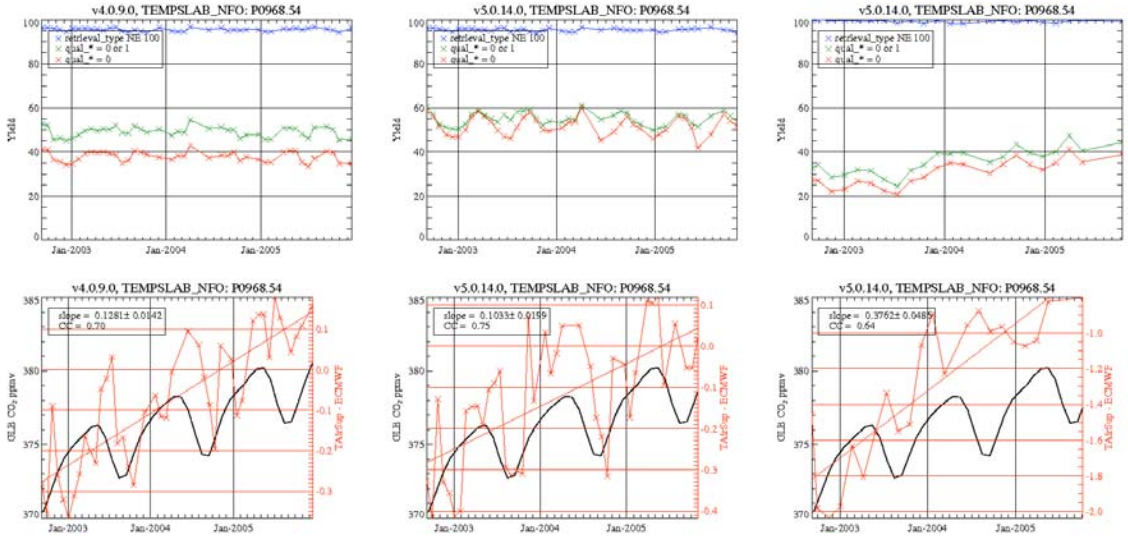

Figure 61 The top panels show the yield for the air temperature retrievals at ~968 mb over non-frozen ocean for V4.0.9.0, V5.0.14.0 AIRS/AMSU, and V5.0.14.0 AIRS-Only (from left to right). The bottom panels show the bias with respect ECMWF for the 3 versions.

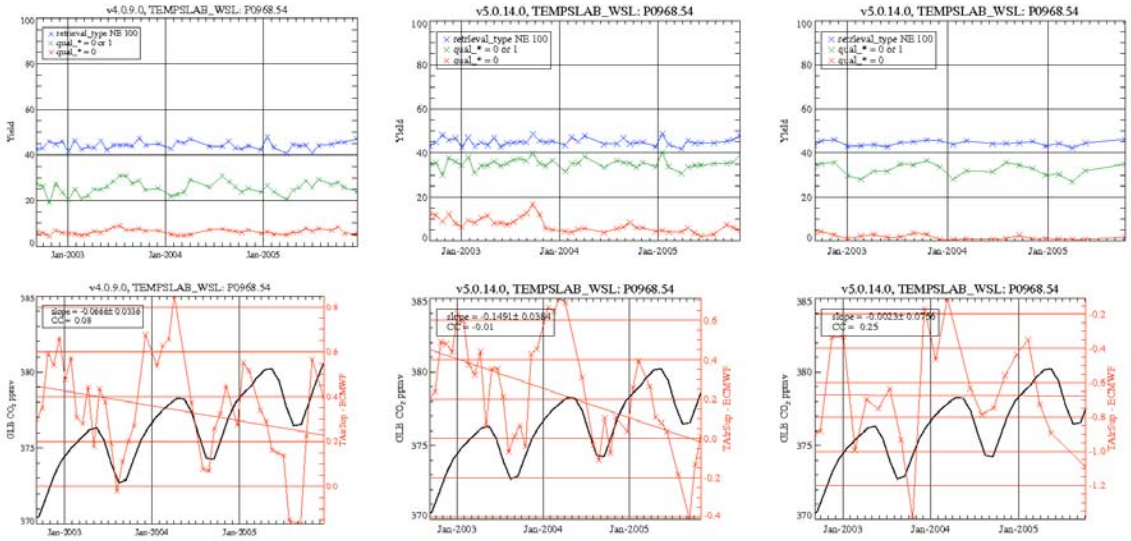

Figure 62 The top panels show the yield for the air temperature retrievals at ~968 mb over warm season land for V4.0.9.0, V5.0.14.0 AIRS/AMSU, and V5.0.14.0 AIRS-Only (from left to right). The bottom panels show the bias with respect ECMWF for the 3 versions.

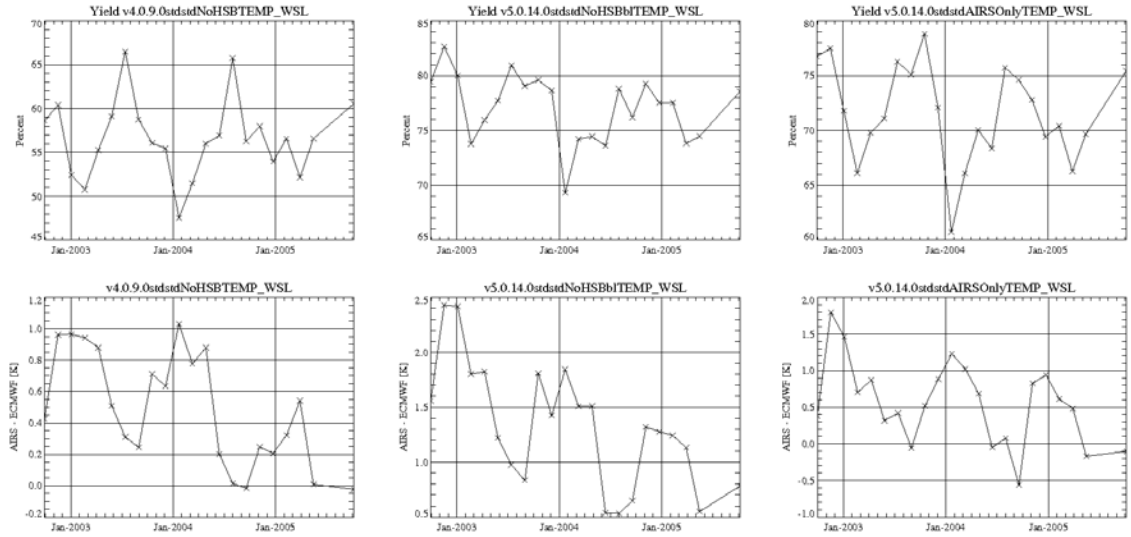

Figure 63 The top panels show a time series of the surface temperature yield over non frozen ocean for V4.0.9.0, V5.0.14.0 AIRS/AMSU, and V5.0.14.0 AIRS-Only (from left to right). The bottom panels show a time series of the difference between the AIRS and the ECMWF land surface temperatures.

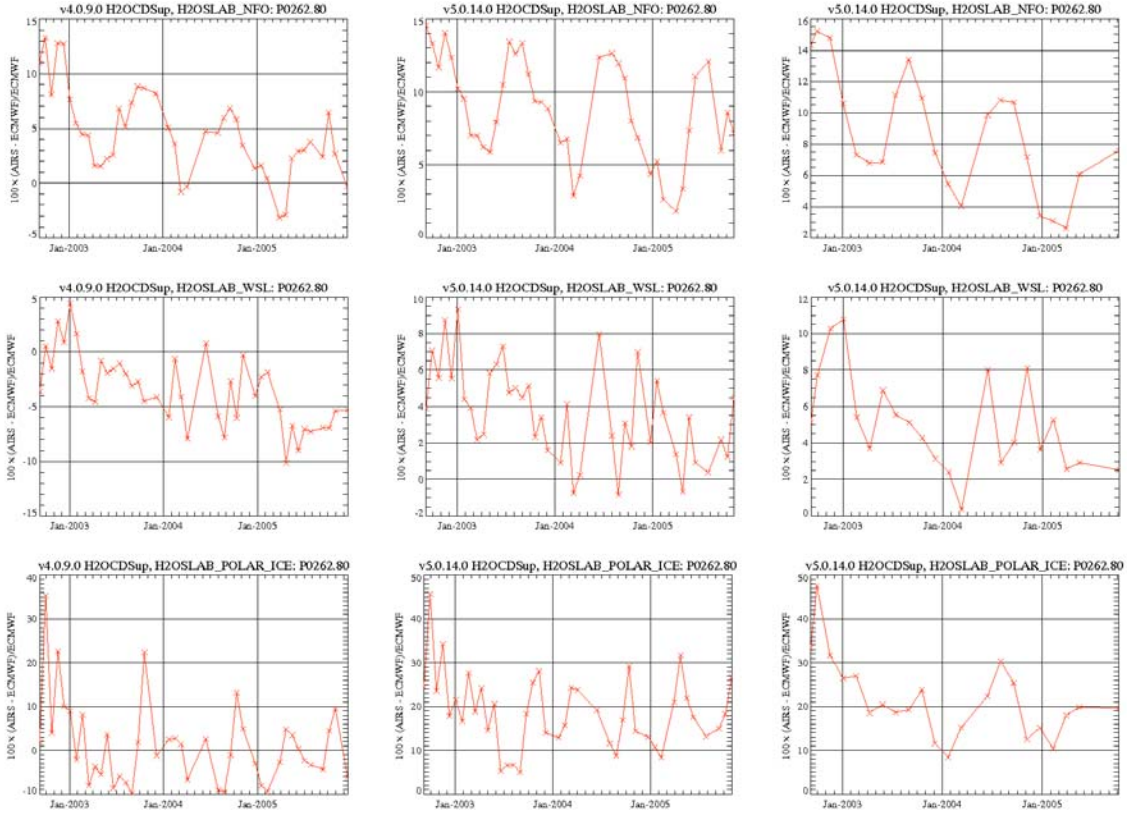

Figure 64 The top row of panels shows trends in the percent error with respect to ECMWF for NFO using V4, V5 AIRS/AMSU, and V5 AIRS-Only. The middle panels show similar plots for WSL, and the bottom panels show the trend for polar ice.

5. Unrealistic Geophysical Conditions

The AIRS retrieval systems often provide “climatologies” that are non-physical, though they may be within reasonable bounds. Two such cases have been identified in V4.0 data: very warm surfaces, and ice-covered surfaces. These and other such problematic situations must be addressed in subsequent data releases. The following sections describe the unrealistic geophysical conditions by PGE version.

5.1. V4.0

Several pathological geophysical states have been identified in V4.0 data. The testing for these cases will involve repeating the analysis that identified the anomalies to verify improvement.

5.1.1. Very Warm Surfaces

The AIRS retrieved fields occasionally show unrealistic conditions over very warm, apparently unvegetated surfaces. Specific locations include the Sahara, Kalahari and Namibia deserts, the Arabian Peninsula, central Australia and the American southwest, all during the warm season. Problems known to occur in these regions are:

- Unrealistically high total ozone.
- Excessive occurrence of low clouds. These clouds are not corroborated by Vis/NIR observations but have a seasonal variation.
- Unrealistically high total water vapor.

5.1.2. Ice-covered Surfaces

Ice-covered surfaces have the following known problems:

- Significant differences in Level 3 Land Surface Temperature compared to MODIS; this may be related to Level 3 quality flagging or MODIS calibration (Reported by R. Knutsen, AGU, AMS, & the March 2006 AIRS Science Team meeting and S. Broberg).
- Anomalously high ozone.
- Quartz emissivities from ice .
- Dry bias relative to ECMWF in summer over Greenland and Antarctica; This may be a problem with ECMWF.

5.1.3. Other known pathologies in V4

- Quartz emissivities over dense tropical forests such as the Congo.
- Upper tropospheric dry bias (Reported in Tobin et al. 2006).
- High total water vapor in microwave-only retrievals in cold air outbreaks (Reported in Fetzer et al. 2006).
- The regression temperature often has a discontinuity just above the surface. For many cases a 40K jump within 30 mb was observed. This discontinuity cannot be fixed by the final algorithm since it uses thick trapezoids to modify temperature profiles. This may be related to the very warm scenes described in Section 4.1.1.
- Cloud cleared radiances are quite often colder than the warmest of the 3 by 3 observed radiances in V4.4.0. This should happen only occasionally in theory. The difference is quite large (-5K) in some cases. This may explain the large difference between MODIS and AIRS ice-covered surface skin temperature differences (See item number 1 in Section 4.1.2).

6. PGE Changes in V5

6.1. New Features

- L3 quant
- trace gases
    - o SO_2 flagging
- dust flagging
- CO retrieval
- CH_4 retrieval
- IR-only capability
- Error estimates
- Linux capability
- Reduced file size

6.2. Improvements in L2 algorithm theoretical basis

- Non-LTE in RTA
- Replaced ad-hoc “model error” with estimate of error due to CO2
- CO_2 climatology first guess
- Reduced bias tuning in IR
- MW tuned to IR
- Channel list changes to reduce impact of land surface emissivity on cloud clearing

7. References

Divakarla, M., C. Barnet, M. D. Goldberg, L. McMillin, E. S. Maddy, W. W. Wolf, L. Zhou, and X. Liu (2006), Validation of Atmospheric Infrared Sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts, J. Geophys. Res., 111, doi:10.1029/2005JD006116.

Fetzer, E. J. (2006), Preface to special section: Validation of Atmospheric Infrared Sounder Observations, J. Geophys. Res., 111, D09S01, doi:10.1029/2005JD007020.

Fetzer, E. J., H. H. Aumann, F. Chen, L. C. Chen, S. Gaiser, D. Hagan, T. Hearty, F. W. Irion, S.-Y. Lee, L. McMillin, E. T. Olsen, H. Revercomb, P. Rosenkranz, D. Staelin, L. L. Strow, J. Susskind, D. Tobin, and J. Zhou (2003), Validation Of AIRS/AMSU/HSB core products for data release version 3.0, August 13, 2003, JPL D-26538, 79 pages, Available at http://disc.gsfc.nasa.gov/AIRS/index.shtml/.

Fetzer, E. J., A. Eldering, E. F. Fishbein, T. Hearty, W. F. Irion and B. Kahn, Validation Of AIRS/AMSU/HSB core products for data release V4.0, March 8, 2005, JPL D-31448, 79 pages, available at http://disc.gsfc.nasa.gov/AIRS/index.shtml/, 2005.

Gettelman, A., E. M. Weinstock, E. J. Fetzer, F. W. Irion, A. Eldering, E. C. Richard, K. H. Rosenlof, T. L. Thompson, J. V. Pittman, C. R. Webster and R. L. Herman, 2004: Validation of satellite data in the upper troposphere and lower stratosphere with in-situ aircraft instruments. Geophys. Res. Lett., vol. 31, L22107, doi:10.1029/2004GL020730.

Gettelman, A., V. P. Walden, L. M. Miloshevich, W. L. Roth, and B. Halter (2006), Relative humidity over Antarctica from radiosondes, satellites, and a general circulation model, J. Geophys. Res., 111, D09S13, doi:10.1029/2005JD006636.

Goldberg, M. D., Y. Qu, L. M. McMillin, W. Wolf, L. Zhou and M. Divakarla (2003), AIRS near-real-time products and algorithms in support of operational numerical weather prediction, IEEE. Trans. Geosci. Remote Sens., 41 (2), 379-389.

Hudson, S. R., & Brandt, R. E., 2005, A look at the Surface-Based Temperature Inversion on the Antarctic Plateau, J. Climate, 18, 1673.

Kahn, B. H, A. Eldering, A. J. Braverman, E. J. Fetzer, J. H. Jiang, E. Fishbein and D. L. Wu, 2006: Towards the characterization of upper tropospheric clouds using AIRS and MLS observations, J. Geophys. Res., submitted.

Kahn, B. H., E. Fishbein, S. L. Nasiri, A. Eldering, E. J. Fetzer, M. J. Garay and S.-Y. Lee, 2006a: The radiative consistency of AIRS and MODIS cloud retrievals, , J. Geophys. Res., submitted.

Levelt, P. F., et al. (2006), Science objectives of the Ozone Monitoring Instrument, IEEE Trans. Geosci. Remote Sens., 44, 1199-1208.

McPeters, R. D., J. A. Logan, G. J. Labow (2003), Ozone Climatological Profiles for Version 8 TOMS and SBUV Retrievals, Eos Trans. AGU, 87 (52), Fall Meet. Suppl., Abstract A21D-0998.

Susskind, J., C. D. Barnet and J. M. Blaisdell (2003), Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds, IEEE Trans. Geosci. and Remote Sensing, 41, 390-409.

Tobin, D.C., H. E. Revercomb, R. O. Knuteson, B. Lesht, L. L. Strow, S. E. Hannon, W. F. Feltz, L. Moy, E. J. Fetzer, and T. Cress (2006), Atmospheric Radiation Measurement site atmospheric state best estimates for Atmospheric Infrared Sounder temperature and water vapor retrieval validation, J. Geophys. Res., 111, doi:10.1029/2005JD006103.

Ziemke, J. R., S. Chandra, and P. K. Bhartia (2001), “Cloud slicing”: A new technique to derive upper tropospheric ozone from satellite measurements, J. Geophys. Res., 106 (D9), 9853-9867.